(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 June 2004 (10.06.2004)

PCT

(10) International Publication Number
WO 2004/047514 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2003/037943

(22) International Filing Date:
25 November 2003 (25.11.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/429,136 25 November 2002 (25.11.2002) US
60/490,234 24 July 2003 (24.07.2003) US

(71) Applicant: **SEQUENOM, INC.** [US/US]; 3595 John Hopkins Court, San Diego, CA 92121 (US).

(72) Inventors: **ROTH, Richard, B.**; 9245 Regents Road, Apartment 409, La Jolla, CA 92037 (US). **NELSON, Matthew, Roberts**; 1250 Calle Prospero, San Marcos, CA 92069 (US). **BRAUN, Andreas**; 3935 Lago Di Grata Circle, San Diego, CA 92130 (US). **KAMMERER, Stefan, M.**; 3825 Elijah Court, Unit 334, San Diego, CA 92130 (US). **RENELAND, Rikard**; 7555 Charmant Drive, #1114, San Diego, CA 92122 (US).

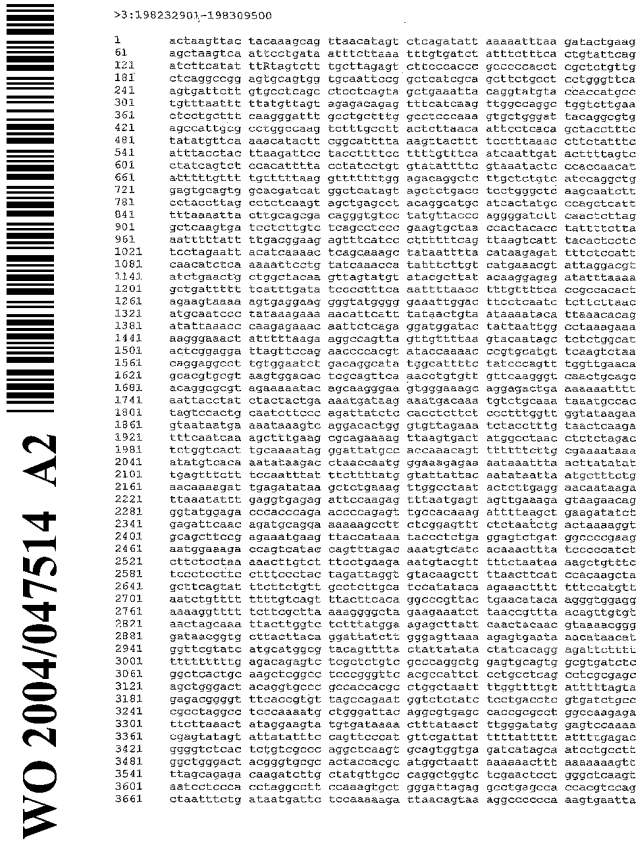
(74) Agents: **GRANT, Bruce, D.** et al.; Morrison & Foerster LLP, 3811 Valley Centre Drive, Sutie 500, San Diego, CA 92130-2332 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Continued on next page]

(54) Title: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS THEREOF



>3:198232901-198309500

```
1       actaagttac tacaaagcag ttaacatagt ctcagatatt aaaaatttaa gatactgaag
61      agctaagtca attcctgata atttcttaaa tttgtgatct atttctttca ctgtattcag
121     atcttcatat ttRtagtctt tgcttagagt ctcccaccc gccccacct cgctctgttg
181     ctcaggccgg agtgcagtgg tgcaattccg gctcatgcca gcttctgcct cctgggttca
241     agtgattctt gtgcctcagc ctcctcagta gctgaaatta caggtatgta ccaccatgcc
301     tgtttaattt ttatgttagt agagacagag tttcatcaag ttggccaggc tggtcttgaa
361     ctcctgtttt caagggattt gcctgctttg gccttccaaa gtgctgggat taccaggcgtg
421     agccattgcg cctggccaag tctttgcctt actcttaaca attcctcaca gctacctttc
481     tatatgttca aaacatactt cggcatttta aagttacttt tcctttaaac cttctatttc
541     atttacctac ttaagattcc taccttttcc ttttgtttca atcaattgat acttttagtc
601     ctatcagtct ccacatttta cctatcctgt gtatattttc gtaaatactc ccaccaacat
661     atttttgttt tgcttttaag gttttttgg agacaggctc ttgctctgtc atccaggctg
721     gagtgcagtg gcacgatcat ggctcatagt agcttctgacc tcctgggctc aagcaatctt
781     cctacctag cctctcaagt agctgagcct acaggcatgc atcactatgc ccagctcatt
841     tttaaattta ctgcagcga caggtgtcc tatgttaccc agggatcttt caactcttag
901     gctcaagtga tcctcttgtc tcagcctccc gaagtgctaa ccactacacc tatttctta
961     aatttttatt ttgacggaag agtttcatcc ctttttcag ttaagtcatt tacactcctc
1021    tcctagaatt acatcaaaac tcagcaaagc tataatttta cataagagat tttctccatt
1081    caacatctca aaaatccctg tatcaaacca tatttcttgt catgaaacgt attaggacgt
1141    atctgaactg ctggctacaa gttagtatgt atacgcttat acaaggagag atatttaaaa
1201    gctgattttt tcatttgata tcccctttca aattttaacc tttgttttca ccgccacact
1261    agaagtaaaa agtgaggaag gggtatgggg gaaattggac ttcctccaatc tcttcttaac
1321    atgcaatccc tatagagaaa aacattcatt tatactgtta atagaatacA ttaaacacag
1381    atattaaacc caagagaaac aattctcaga ggatggatac tattaattgg cctaaagaaa
1441    aagggaaact atttttaaga aggccagtta gttgttttaa gtacaatagc tctctggcat
1501    actcggagga ttagttccag aaccccacgt ataccaaaac ccgtgcatgt tcaagtctaa
1561    caggaggcct tgtggaatct gacaggcata tggcattttc tatcccagtt tggttgaaca
1621    gcacgtgcgt aagtggacac tcgcagttca aacctgtgtt gttcaagggt caactgcagc
1681    acaggcgcgt agaaaaatac agcaagggaa gtgggaaagc aggagactga aaaaaatttt
1741    aattacctat ctactactga aaatgataag aaatgacaaa tgtctgcaaa taaatgccac
1801    tagtccactg caatcttccc agattatctc cacctcttct ccctttggtt ggtataagaa
1861    gtaataatga aaataaagtc aggacactgg gtgttagaaa tctacctttg taactcaaga
1921    tttcaatcaa agctttgaag cgcagaaaag ttaagtgact atggcctaac ctctctagac
1981    tctggtcact tgcaaaatag ggattatgcc accaaacagt ttttttcttg cgaaaataaa
2041    atatgtcaca aatataagac ctaaccaatg ggaaagagaa aataaattta acttatatat
2101    tgagtttctt tccaatttat ttcttttatg gtattattac aatataatta atgctttctg
2161    aacaaagat tgagatataa gctctgaaag ttggcctaat actcttgagg aacaataaga
2221    ttaaatattt gaggtgagag attccaagag tttaatgagt agttgaaaga gtaagaacag
2281    ggtatggaga cccacccaga acccagagt tgccacaaag attttaagct gaagatatct
2341    gagattcaac agatgcagga aaaaagcctt ctcggagttt ctctaatctg actaaaaggt
2401    gcagctttcg agaaatgaag ttaccataaa taccctctga ggagtctgat ggccccgaag
2461    aatggaaaga ccagtcatac cagtttagac aaatgtcatc acaaacttta tccccatct
2521    cttctcctaa aaacttgtct ttcctgaaga aatgtacgtt tttctaataa aagctgtttc
2581    tccctccttc ctttccctac tagattaggt gtacaagctt ttaacttcat ccacaagcta
2641    gcttcagtat ttcttctgtt gcctcttgca tccatataca agaaactttt ttttccatgtt
2701    aatctgtttt ttttgtcagt ttacttcaca ggcccgttac tgaacataca agggtggagg
2761    aaaaggtttt tcttcgctta aaaggggcta gaagaaatct taaccgttta acagttgtgt
2821    aactagcaaa ttacttggtc tctttatgga agagcttatt caactacaac gtaaaacggg
2881    gataacggtg cttacttaca ggattatctt gggagttaaa agagtgaata aacataacat
2941    ggttcgtatc atgcatggcg tacagttttA ctattatata ctatcacagg agatctttt
3001    ttttttttg agacagagtc tcgctctgtc gcccaggctg gagtgcagtg gcgtgatctc
3061    ggctcactgc aagctcggcc tcccgggttc acgccattct cctgcctcag cctcgcgagc
3121    agctgggact acaggtgccc gccaccacgc ctggctaatt ttgttttgt attttagta
3181    gagacggggt ttcaccgtgt tagccagaat ggtctctatc tcctgacctc gtgatctgcc
3241    cgcctaggcc tcccaaaatg ctgggattac aggcgtgagc caccgcgcct ggccaagaga
3301    ttcttaaact ataggaagta tgtgataaaa ctttataact ttgggatatg gagtccaaaa
3361    cgagtatagt attatatttc cagttcccat gttcgattat ttttatttt attttgagac
3421    ggggtctcac tctgtcgccc aggctcaagt gcagtggtga gatcatagca atcctgcctt
3481    ggctgggact acgggtgcgc actaccacgc atggctaatt aaaaaacttt aaaaaaagtc
3541    ttagcagaga caagatcttg ctatgttgcc caggctggtc tcgaactcct gggctcaagt
3601    aatcctccca cctaggcttt ccaaagtgct gggattagag gcctgagcca ccacgtccag
3661    ctaatttctg ataatgattc tccaaaaaga ttaacagtaa aggccccca aagtgaatta
```

(57) Abstract: Provided herein are methods for identifying risk of breast cancer in a subject and/or a subject at risk of breast cancer, reagents and kits for carrying out the methods, methods for identifying candidate therapeutics for treating breast cancer, and therapeutic methods for treating breast cancer in a subject. These embodiments are based upon an analysis of polymorphic variations in nucleotide sequences within the human genome.

WO 2004/047514 A2

Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS THEREOF

Field of the Invention

[0001] The invention relates to genetic methods for identifying risk of breast cancer and treatments that specifically target the disease.

Background

[0002] Breast cancer is the third most common cancer, and the most common cancer in women, as well as a cause of disability, psychological trauma, and economic loss. Breast cancer is the second most common cause of cancer death in women in the United States, in particular for women between the ages of 15 and 54, and the leading cause of cancer-related death (Forbes, Seminars in Oncology, vol.24(1), Suppl 1, 1997: pp.S1-20-S1-35). Indirect effects of the disease also contribute to the mortality from breast cancer including consequences of advanced disease, such as metastases to the bone or brain. Complications arising from bone marrow suppression, radiation fibrosis and neutropenic sepsis, collateral effects from therapeutic interventions, such as surgery, radiation, chemotherapy, or bone marrow transplantation-also contribute to the morbidity and mortality from this disease.

[0003] While the pathogenesis of breast cancer is unclear, transformation of normal breast epithelium to a malignant phenotype may be the result of genetic factors, especially in women under thirty (Miki, *et al.*, Science, 266: 66-71 (1994)). However, it is likely that other, non-genetic factors also have a significant effect on the etiology of the disease. Regardless of its origin, breast cancer morbidity increases significantly if it is not detected early in its progression. Thus, considerable efforts have focused on the elucidation of early cellular events surrounding transformation in breast tissue. Such efforts have led to the identification of several potential breast cancer markers. For example, alleles of the *BRCA1* and *BRCA2* genes have been linked to hereditary and early-onset breast cancer (Wooster, *et al.*, Science, 265: 2088-2090 (1994)). However, *BRCA1* is limited as a cancer marker because *BRCA1* mutations fail to account for the majority of breast cancers (Ford, *et al.*, British J. Cancer, 72: 805-812 (1995)). Similarly, the *BRCA2* gene, which has been linked to forms of hereditary breast cancer, accounts for only a small portion of total breast cancer cases.

Summary

[0004] It has been discovered that certain polymorphic variations in human genomic DNA are associated with the occurrence of breast cancer. In particular, polymorphic variants in loci containing

DLG1, KIAA0783, DPF3 and *CENPC1* regions in human genomic DNA have been associated with risk of breast cancer.

[0005] Thus, featured herein are methods for identifying a subject at risk of breast cancer and/or a risk of breast cancer in a subject, which comprises detecting the presence or absence of one or more polymorphic variations associated with breast cancer in genomic regions described herein in a human nucleic acid sample. In an embodiment, two or more polymorphic variations are detected in two or more regions selected from the group consisting of *DLG1, KIAA0783, DPF3* and *CENPC1*. In certain embodiments, 3 or fewer, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or fewer polymorphic variants are detected.

[0006] Also featured are nucleic acids that include one or more polymorphic variations associated with the occurrence of breast cancer, as well as polypeptides encoded by these nucleic acids. Further, provided is a method for identifying a subject at risk of breast cancer and then prescribing to the subject a breast cancer detection procedure, prevention procedure and/or a treatment procedure. In addition, provided are methods for identifying candidate therapeutic molecules for treating breast cancer and related disorders, as well as methods for treating breast cancer in a subject by diagnosing breast cancer in the subject and treating the subject with a suitable treatment, such as administering a therapeutic molecule.

[0007] Also provided are compositions comprising a breast cancer cell and/or *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid with a RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid designed from a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence. In an embodiment, the nucleic acid is designed from a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence that includes one or more breast cancer associated polymorphic variations, and in some instances, specifically interacts with such a nucleotide sequence. Further, provided are arrays of nucleic acids bound to a solid surface, in which one or more nucleic acid molecules of the array have a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence, or a fragment or substantially identical nucleic acid thereof, or a complementary nucleic acid of the foregoing. Featured also are compositions comprising a breast cancer cell and/or a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide, with an antibody that specifically binds to the polypeptide. In an embodiment, the antibody specifically binds to an epitope in the polypeptide that includes a non-synonymous amino acid modification associated with breast cancer (e.g., results in an amino acid substitution in the encoded polypeptide associated with breast cancer). In certain embodiments, the antibody specifically binds to an epitope that comprises a glutamine at amino acid position 278 in SEQ ID NO: 9 of a *DLG1* polypeptide or a glycine at amino acid position 389 in SEQ ID NO: 12 of a *CENPC1* polypeptide.

Brief Description of the Figures

[0008] Figures 1A-1T show a genomic nucleotide sequence for an *DLG1* region. The genomic nucleotide sequence is set forth in SEQ ID NO: 1. The following nucleotide representations are used

throughout: "A" or "a" is adenosine, adenine, or adenylic acid; "C" or "c" is cytidine, cytosine, or cytidylic acid; "G" or "g" is guanosine, guanine, or guanylic acid; "T" or "t" is thymidine, thymine, or thymidylic acid; and "I" or "i" is inosine, hypoxanthine, or inosinic acid. Exons are indicated in italicized lower case type, introns are depicted in normal text lower case type, and polymorphic sites are depicted in bold upper case type. SNPs are designated by the following convention: "R" represents A or G, "M" represents A or C; "W" represents A or T; "Y" represents C or T; "S" represents C or G; "K" represents G or T; "V" represents A, C or G; "H" represents A, C, or T; "D" represents A, G, or T; "B" represents C, G, or T; and "N" represents A, G, C, or T.

[0009] Figures 2A-2Z show a genomic nucleotide sequence of a *KIAA0783* region. The genomic nucleotide sequence is set forth in SEQ ID NO: 2.

[0010] Figures 3A-3X show a genomic nucleotide sequence of a *DPF3* region. The genomic nucleotide sequence is set forth in SEQ ID NO: 3.

[0011] Figures 4A-4Y show a genomic nucleotide sequence of a *CENPC1* region. The genomic nucleotide sequence is set forth in SEQ ID NO: 4.

[0012] Figure 5 shows a coding nucleotide sequence (cDNA) for *DLG1*. The nucleotide sequence is set forth in SEQ ID NO: 5.

[0013] Figure 6 shows a coding nucleotide sequence (cDNA) for *KIAA0783*. The nucleotide sequence is set forth in SEQ ID NO: 6.

[0014] Figure 7 shows a coding nucleotide sequence (cDNA) for *DPF3*. The nucleotide sequence is set forth in SEQ ID NO: 7.

[0015] Figure 8 shows a coding nucleotide sequence (cDNA) for *CENPC1*. The nucleotide sequence is set forth in SEQ ID NO: 8.

[0016] Figure 9 shows an amino acid sequence for a *DLG1* polypeptide, which is set forth in SEQ ID NO: 9.

[0017] Figure 10 shows an amino acid sequence for a *KIAA0783* polypeptide, which is set forth in SEQ ID NO: 10.

[0018] Figure 11 shows an amino acid sequence for a *DPF3* polypeptide, which is set forth in SEQ ID NO: 11.

[0019] Figure 12 shows an amino acid sequence for a *CENPC1* polypeptide, which is set forth in SEQ ID NO: 12.

[0020] Figures 13-16 show proximal SNPs in *DLG1, KIAA0783, DPF3* and *CENPC1* loci in genomic DNA. The position of each SNP on the chromosome is shown on the x-axis and the y-axis provides the negative logarithm of the p-value comparing the estimated allele to that of the control group. Also shown in the figure are exons and introns of the genes in the approximate chromosomal positions. The figure indicates that polymorphic variants associated with breast cancer are in linkage disequilibrium in the following regions: the region spanning positions 7938-59808 in SEQ ID NO: 1;

the region spanning positions 10511-98107 in SEQ ID NO: 2; the region spanning positions 160-72752 in SEQ ID NO: 3; and the region spanning positions 196-74909 in SEQ ID NO: 4.

Detailed Description

[0021] It has been discovered that polymorphic variations in the *DLG1, KIAA0783, DPF3* and *CENPC1* regions described herein are associated with an increased risk of breast cancer.

[0022] The gene *DLG1* (discs, large homolog 1 (Drosophila)) is also referenced as synapse-associated protein 97, hdlg, SAP97. *DLG1* has been mapped to chromosomal position 3-q29. In Drosophila more than 50 genes have been identified that lead to loss of cell proliferation control, indicating that they are tumor suppressor genes. Many of these genes have been cloned and sequenced, and most have clear mammalian homologs. The Drosophila 'discs large' tumor suppressor protein, Dlg, is the prototype of a family of proteins termed MAGUKs (membrane-associated guanylate kinase homologs). MAGUKs are localized at the membrane-cytoskeleton interface, usually at cell-cell junctions, where they appear to have both structural and signaling roles. They contain several distinct domains, including a modified guanylate kinase domain, an SH3 motif, and 1 or 3 copies of the DHR (GLGF/PDZ) domain. Recessive lethal mutations in the 'discs large' tumor suppressor gene interfere with the formation of septate junctions (thought to be the arthropod equivalent of tight junctions) between epithelial cells, and they also cause neoplastic overgrowth of imaginal discs, suggesting a role for cell junctions in proliferation control.

[0023] The gene *KIAA0783* also is known as PHF14 and PHD finger protein 14. *KIAA0783* has been mapped to chromosomal position 7p21.3. The protein encoded by this gene is a novel gene with unknown function. Being a zinc finger protein, it likely a transcription factor.

[0024] The gene *DPF3* (D4, zinc and double PHD fingers, family 3) also is known as CERD4, cer-d4, FLJ14079, and 2810403B03Rik. DPF3 is a Rho family guanine-nucleotide exchange factor. *DPF3* has been mapped to chromosomal position 14q24.3-q31.1.

[0025] The gene *CENPC1* (centromere protein C1) also is known as Centromere autoantigen C1. CENPC1 has been mapped to chromosomal position 4q12-q13.3. *CENPC1* is a centromere autoantigen and a component of the inner kinetochore plate. The protein is required for maintaining proper kinetochore size and a timely transition to anaphase. A putative pseudogene exists on chromosome 12.

Breast Cancer and Sample Selection

[0026] Breast cancer is typically described as the uncontrolled growth of malignant breast tissue. Breast cancers arise most commonly in the lining of the milk ducts of the breast (ductal carcinoma), or in the lobules where breast milk is produced (lobular carcinoma). Other forms of breast cancer include Inflammatory Breast Cancer and Recurrent Breast Cancer. Inflammatory breast cancer is a

rare, but very serious, aggressive type of breast cancer. The breast may look red and feel warm with ridges, welts, or hives on the breast; or the skin may look wrinkled. It is sometimes misdiagnosed as a simple infection. Recurrent disease means that the cancer has come back after it has been treated. It may come back in the breast, in the soft tissues of the chest (the chest wall), or in another part of the body.

[0027] As used herein, the term "breast cancer" refers to a condition characterized by anomalous rapid proliferation of abnormal cells in one or both breasts of a subject. The abnormal cells often are referred to as "neoplastic cells," which are transformed cells that can form a solid tumor. The term "tumor" refers to an abnormal mass or population of cells (*i.e.* two or more cells) that result from excessive or abnormal cell division, whether malignant or benign, and pre-cancerous and cancerous cells. Malignant tumors are distinguished from benign growths or tumors in that, in addition to uncontrolled cellular proliferation, they can invade surrounding tissues and can metastasize. In breast cancer, neoplastic cells may be identified in one or both breasts only and not in another tissue or organ, in one or both breasts and one or more adjacent tissues or organs (*e.g.* lymph node), or in a breast and one or more non-adjacent tissues or organs to which the breast cancer cells have metastasized.

[0028] The term "invasion" as used herein refers to the spread of cancerous cells to adjacent surrounding tissues. The term "invasion" often is used synonymously with the term "metastasis," which as used herein refers to a process in which cancer cells travel from one organ or tissue to another non-adjacent organ or tissue. Cancer cells in the breast(s) can spread to tissues and organs of a subject, and conversely, cancer cells from other organs or tissue can invade or metastasize to a breast. Cancerous cells from the breast(s) may invade or metastasize to any other organ or tissue of the body. Breast cancer cells often invade lymph node cells and/or metastasize to the liver, brain and/or bone and spread cancer in these tissues and organs. Breast cancers can spread to other organs and tissues and cause lung cancer, prostate cancer, colon cancer, ovarian cancer, cervical cancer, gastrointestinal cancer, pancreatic cancer, glioblastoma, bladder cancer, hepatoma, colorectal cancer, uterine cervical cancer, endometrial carcinoma, salivary gland carcinoma, kidney cancer, vulval cancer, thyroid cancer, hepatic carcinoma, skin cancer, melanoma, ovarian cancer, neuroblastoma, myeloma, various types of head and neck cancer, acute lymphoblastic leukemia, acute myeloid leukemia, Ewing sarcoma and peripheral neuroepithelioma, and other carcinomas, lymphomas, blastomas, sarcomas, and leukemias.

[0029] Breast cancers arise most commonly in the lining of the milk ducts of the breast (ductal carcinoma), or in the lobules where breast milk is produced (lobular carcinoma). Other forms of breast cancer include Inflammatory Breast Cancer and Recurrent Breast Cancer. Inflammatory Breast Cancer is a rare, but very serious, aggressive type of breast cancer. The breast may look red and feel warm with ridges, welts, or hives on the breast; or the skin may look wrinkled. It is sometimes misdiagnosed as a simple infection. Recurrent disease means that the cancer has come back after it

has been treated. It may come back in the breast, in the soft tissues of the chest (the chest wall), or in another part of the body. As used herein, the term "breast cancer" may include both Inflammatory Breast Cancer and Recurrent Breast Cancer.

[0030] In an effort to detect breast cancer as early as possible, regular physical exams and screening mammograms often are prescribed and conducted. A diagnostic mammogram often is performed to evaluate a breast complaint or abnormality detected by physical exam or routine screening mammography. If an abnormality seen with diagnostic mammography is suspicious, additional breast imaging (with exams such as ultrasound) or a biopsy may be ordered. A biopsy followed by pathological (microscopic) analysis is a definitive way to determine whether a subject has breast cancer. Excised breast cancer samples often are subjected to the following analyses: diagnosis of the breast tumor and confirmation of its malignancy; maximum tumor thickness; assessment of completeness of excision of invasive and *in situ* components and microscopic measurements of the shortest extent of clearance; level of invasion; presence and extent of regression; presence and extent of ulceration; histological type and special variants; pre-existing lesion; mitotic rate; vascular invasion; neurotropism; cell type; tumor lymphocyte infiltration; and growth phase.

[0031] The stage of a breast cancer can be classified as a range of stages from Stage 0 to Stage IV based on its size and the extent to which it has spread. The following table summarizes the stages:

Table A

Stage	Tumor Size	Lymph Node Involvement	Metastasis (Spread)
I	Less than 2 cm	No	No
II	Between 2-5 cm	No or in same side of breast	No
III	More than 5 cm	Yes, on same side of breast	No
IV	Not applicable	Not applicable	Yes

[0032] Stage 0 cancer is a contained cancer that has not spread beyond the breast ductal system. Fifteen to twenty percent of breast cancers detected by clinical examinations or testing are in Stage 0 (the earliest form of breast cancer). Two types of Stage 0 cancer are lobular carcinoma in situ (LCIS) and ductal carcinoma in situ (DCIS). LCIS indicates high risk for breast cancer. Many physicians do not classify LCIS as a malignancy and often encounter LCIS by chance on breast biopsy while investigating another area of concern. While the microscopic features of LCIS are abnormal and are similar to malignancy, LCIS does not behave as a cancer (and therefore is not treated as a cancer). LCIS is merely a marker for a significantly increased risk of cancer anywhere in the breast. However, bilateral simple mastectomy may be occasionally performed if LCIS patients have a strong family

history of breast cancer. In DCIS the cancer cells are confined to milk ducts in the breast and have not spread into the fatty breast tissue or to any other part of the body (such as the lymph nodes). DCIS may be detected on mammogram as tiny specks of calcium (known as microcalcifications) 80% of the time. Less commonly DCIS can present itself as a mass with calcifications (15% of the time); and even less likely as a mass without calcifications (<5% of the time). A breast biopsy is used to confirm DCIS. A standard DCIS treatment is breast-conserving therapy (BCT), which is lumpectomy followed by radiation treatment or mastectomy. To date, DCIS patients have chosen equally among lumpectomy and mastectomy as their treatment option, though specific cases may sometimes favor lumpectomy over mastectomy or vice versa.

[0033] In Stage I, the primary (original) cancer is 2 cm or less in diameter and has not spread to the lymph nodes. In Stage IIA, the primary tumor is between 2 and 5 cm in diameter and has not spread to the lymph nodes. In Stage IIB, the primary tumor is between 2 and 5 cm in diameter and has spread to the axillary (underarm) lymph nodes; or the primary tumor is over 5 cm and has not spread to the lymph nodes. In Stage IIIA, the primary breast cancer of any kind that has spread to the axillary (underarm) lymph nodes and to axillary tissues. In Stage IIIB, the primary breast cancer is any size, has attached itself to the chest wall, and has spread to the pectoral (chest) lymph nodes. In Stage IV, the primary cancer has spread out of the breast to other parts of the body (such as bone, lung, liver, brain). The treatment of Stage IV breast cancer focuses on extending survival time and relieving symptoms.

[0034] Based in part upon selection criteria set forth above, individuals having breast cancer can be selected for genetic studies. Also, individuals having no history of cancer or breast cancer often are selected for genetic studies. Other selection criteria can include: a tissue or fluid sample is derived from an individual characterized as Caucasian; the sample was derived from an individual of German paternal and maternal descent; the database included relevant phenotype information for the individual; case samples were derived from individuals diagnosed with breast cancer; control samples were derived from individuals free of cancer and no family history of breast cancer; and sufficient genomic DNA was extracted from each blood sample for all allelotyping and genotyping reactions performed during the study. Phenotype information included pre- or post-menopausal, familial predisposition, country or origin of mother and father, diagnosis with breast cancer (date of primary diagnosis, age of individual as of primary diagnosis, grade or stage of development, occurrence of metastases, *e.g.*, lymph node metastases, organ metastases), condition of body tissue (skin tissue, breast tissue, ovary tissue, peritoneum tissue and myometrium), method of treatment (surgery, chemotherapy, hormone therapy, radiation therapy).

[0035] Provided herein is a set of blood samples and a set of corresponding nucleic acid samples isolated from the blood samples, where the blood samples are donated from individuals diagnosed with breast cancer. The sample set often includes blood samples or nucleic acid samples from 100 or more, 150 or more, or 200 or more individuals having breast cancer, and sometimes from 250 or

more, 300 or more, 400 or more, or 500 or more individuals. The individuals can have parents from any place of origin, and in an embodiment, the set of samples are extracted from individuals of German paternal and German maternal ancestry. The samples in each set may be selected based upon five or more criteria and/or phenotypes set forth above.

Polymorphic Variants Associated with Breast Cancer

[0036] A genetic analysis provided herein linked breast cancer with polymorphic variants in the *DLG1, KIAA0783, DPF3* and *CENPC1* regions of the human genome disclosed herein. As used herein, the term "polymorphic site" refers to a region in a nucleic acid at which two or more alternative nucleotide sequences are observed in a significant number of nucleic acid samples from a population of individuals. A polymorphic site may be a nucleotide sequence of two or more nucleotides, an inserted nucleotide or nucleotide sequence, a deleted nucleotide or nucleotide sequence, or a microsatellite, for example. A polymorphic site that is two or more nucleotides in length may be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more, 20 or more, 30 or more, 50 or more, 75 or more, 100 or more, 500 or more, or about 1000 nucleotides in length, where all or some of the nucleotide sequences differ within the region. A polymorphic site is often one nucleotide in length, which is referred to herein as a "single nucleotide polymorphism" or a "SNP."

[0037] Where there are two, three, or four alternative nucleotide sequences at a polymorphic site, each nucleotide sequence is referred to as a "polymorphic variant" or "nucleic acid variant." Where two polymorphic variants exist, for example, the polymorphic variant represented in a minority of samples from a population is sometimes referred to as a "minor allele" and the polymorphic variant that is more prevalently represented is sometimes referred to as a "major allele." Many organisms possess a copy of each chromosome (*e.g.*, humans), and those individuals who possess two major alleles or two minor alleles are often referred to as being "homozygous" with respect to the polymorphism, and those individuals who possess one major allele and one minor allele are normally referred to as being "heterozygous" with respect to the polymorphism. Individuals who are homozygous with respect to one allele are sometimes predisposed to a different phenotype as compared to individuals who are heterozygous or homozygous with respect to another allele.

[0038] Furthermore, a genotype or polymorphic variant may be expressed in terms of a "haplotype," which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0039] As used herein, the term "phenotype" refers to a trait which can be compared between individuals, such as presence or absence of a condition, a visually observable difference in appearance between individuals, metabolic variations, physiological variations, variations in the function of biological molecules, and the like. An example of a phenotype is occurrence of breast cancer.

[0040] Researchers sometimes report a polymorphic variant in a database without determining whether the variant is represented in a significant fraction of a population. Because a subset of these reported polymorphic variants are not represented in a statistically significant portion of the population, some of them are sequencing errors and/or not biologically relevant. Thus, it is often not known whether a reported polymorphic variant is statistically significant or biologically relevant until the presence of the variant is detected in a population of individuals and the frequency of the variant is determined. Methods for detecting a polymorphic variant in a population are described herein, specifically in Example 2. A polymorphic variant is statistically significant and often biologically relevant if it is represented in 5% or more of a population, sometimes 10% or more, 15% or more, or 20% or more of a population, and often 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, or 50% or more of a population.

[0041] A polymorphic variant may be detected on either or both strands of a double-stranded nucleic acid. For example, a thymine at a particular position in SEQ ID NO: 1 can be reported as an adenine from the complementary strand. Also, a polymorphic variant may be located within an intron or exon of a gene or within a portion of a regulatory region such as a promoter, a 5′ untranslated region (UTR), a 3′ UTR, and in DNA (*e.g.*, genomic DNA (gDNA) and complementary DNA (cDNA)), RNA (*e.g.*, mRNA, tRNA, and rRNA), or a polypeptide. Polymorphic variations may or may not result in detectable differences in gene expression, polypeptide structure, or polypeptide function.

[0042] In the genetic analysis that associated breast cancer with the polymorphic variants described hereafter, samples from individuals having breast cancer and individuals not having cancer were allelotyped and genotyped. The term "genotyped" as used herein refers to a process for determining a genotype of one or more individuals, where a "genotype" is a representation of one or more polymorphic variants in a population. Genotypes may be expressed in terms of a "haplotype," which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0043] It was determined that polymorphic variations associated with an increased risk of breast cancer existed in *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequences. Polymorphic variants

in and around the *DLG1, KIAA0783, DPF3* and *CENPC1* loci were tested for association with breast cancer. In the *DLG1* locus, these included polymorphic variants at positions in SEQ ID NO: 1 selected from the group consisting of 133, 7938, 8873, 13221, 17288, 25732, 26923, 39977, 41284, 41410, 41477, 41514, 42606, 42742, 59515, 59808, 60265, 67152, 68332, 71128 and 76427. Polymorphic variants in a region spanning positions 7938-59808 in SEQ ID NO: 1 in particular were associated with an increased risk of breast cancer, including polymorphic variants at positions 7938, 26923, 39977 and 59808 in SEQ ID NO: 1. At these positions in SEQ ID NO: 1, a thymine at position 7938, a cytosine at position 26923, a thymine at position 39977 and a thymine at position 59808 in particular were associated with risk of breast cancer. Also, a glutamine at position 278 in SEQ ID NO: 9 in a *DLG1* polypeptide in particular was associated with an increased risk of breast cancer.

[0044] In the *KIAA0783* locus, these included polymorphic variants at positions in SEQ ID NO: 2 selected from the group consisting of 201, 6395, 8558, 9429, 9809, 10072, 10511, 11556, 16857, 16951, 17027, 17177, 17615, 17950, 18329, 18384, 18561, 18579, 18871, 27152, 27306, 28091, 28661, 29011, 29962, 29969, 30085, 31656, 31685, 31749, 45389, 45459, 46647, 49860, 53061, 57308, 61563, 61660, 62212, 67090, 67198, 70071, 70191, 74006, 75600, 85761, 90798, 90883, 91259, 95416, 95446, 96368, 97050, 97362, 97630, 97989 and 98107. Polymorphic variants in a region spanning positions 10511-98107 in SEQ ID NO: 2 in particular were associated with an increased risk of breast cancer, including polymorphic variants at positions 10511, 11556, 17177, 18384, 28661, 31656, 31685, 31749, 45389, 45459, 46647, 49860, 53061, 57308, 61563, 61660, 67090, 67198, 70071, 74006, 75600, 85761, 90798, 90883, 91259, 95416, 95446, 96368, 97362, 97630, 97989 and 98107 in SEQ ID NO: 2. At these positions in SEQ ID NO: 2, a thymine at position 10511, a cytosine at position 11556, a thymine at position 17177, a thymine at position 18384, an adenine at position 28661, an adenine at position 31656, an adenine at position 31685, a guanine at position 31749, a thymine at position 45389, a guanine at position 45459, an adenine at position 46647, a thymine at position 49860, a thymine at position 53061, an adenine at position 57308, a guanine at position 61563, a guanine at position 61660, a guanine at position 67090, a cytosine at position 67198, an adenine at position 70071, a cytosine at position 74006, an adenine at position 75600, a guanine at position 85761, a thymine at poisition 90798, a cytosine at position 90883, an adenine at position 91259, a cytosine at position 95416, a thymine at position 95446, a thymine at position 96368, a thymine at position 97362, an adenine at position 97630, a cytosine at position 97989 and a thymine at position 98107 in particular were associated with increased risk of breast cancer.

[0045] In the *DPF3* locus, these included polymorphic variants at positions in SEQ ID NO: 3 selected from the group consisting of 160, 6053, 9719, 10481, 10676, 17179, 18561, 18658, 18694, 18858, 24582, 24683, 24767, 27402, 28150, 28494, 32003, 35588, 35619, 35856, 36254, 37314, 40033, 40095, 42593, 42799, 43090, 46683, 49774, 51796, 52079, 53857, 53971, 55899, 60682,

61291, 72720, 72752, 85507 and 89751. Polymorphic variants in a region spanning positions 160-72752 in SEQ ID NO: 3 in particular were associated with an increased risk of breast cancer, including polymorphic variants at positions 160, 6053, 18658, 18694, 18858, 24683, 27402, 28494, 32003, 35588, 35856, 40095, 46683, 52079, 53857, 72720 and 72752 in SEQ ID NO: 3. At these positions in SEQ ID NO: 3, an adenine at position 160, a guanine at position 6053, a guanine at position 18658, a guanine at position 18694, a thymine at position 18858, a guanine at position 24683, a guanine at position 27402, a thymine at position 28494, an adenine at position 32003, a cytosine at position 35588, an adenine at position 35856, a guanine at position 40095, an adenine at position 46683, an adenine at position 52079, a cytosine at position 53857, an adenine at position 72720 and a cytosine at position 72752 in particular were associated with an increased risk of breast cancer.

[0046] In the *CENPC1* locus, these included polymorphic variants at positions in SEQ ID NO: 4 selected from the group consisting of 196, 13311, 14486, 14691, 15551, 17702, 17872, 19588, 19910, 20006, 20575, 21092, 22830, 23455, 23716, 23890, 24001, 24995, 27282, 27779, 29099, 31185, 33994, 34942, 35137, 36538, 37139, 37358, 38828, 39469, 40233, 40472, 41679, 41682, 42831, 42976, 44128, 44195, 46769, 47363, 48843, 52574, 52602, 53212, 53781, 54710, 55808, 57987, 58556, 59148, 59286, 60217, 60412, 60753, 60791, 61524, 62543, 62825, 62826, 62857, 63400, 63960, 64307, 64539, 65728, 66000, 66521, 68185, 69643, 74909, 82973, 83039, 85713, 86873, 90293, 91810, 92609, 92884 and 42831. Polymorphic variants in a region spanning positions 196-74909 in SEQ ID NO: 4 in particular were associated with an increased risk of breast cancer, including polymorphic variants at positions 196, 13311, 14486, 19910, 20575, 23716, 23890, 24995, 29099, 33994, 34942, 37139, 40233, 40472,42831, 42976, 44195, 48843, 58556, 59286, 60217, 62826, 62857, 63400, 63960 and 74909 in SEQ ID NO: 4. At these positions in SEQ ID NO: 4, an adenine at position 196, a guanine at position 13311, a thymine at position 14486, a thymine at position 19910, an adenine at position 20575, a guanine at position 23716, a guanine at position 23890, an adenine at position 24995, a cytosine at position 29099, a thymine at position 33994, a thymine at position 34942, a thymine at position 37139, a thymine at position 40233, an adenine at position 40472, a guanine at position 42831, a guanine at position 42976, a thymine at position 44195, a thymine at position 48843, an adenine at position 58556, a guanine at position 59286, an adenine at position 60217, a cytosine at position 62826, a thymine at position 62857, a thymine at position 63400, an adenine at position 63960 and a cytosine at position 74909 in particular were associated with an increased risk of breast cancer. Also, a glycine at position 389 in SEQ ID NO: 12 in a *CENPC1* polypeptide in particular was associated with an increased risk of breast cancer.

Additional Polymorphic Variants Associated with Breast Cancer

[0047] Also provided is a method for identifying polymorphic variants proximal to an incident, founder polymorphic variant associated with breast cancer. Thus, featured herein are methods for identifying a polymorphic variation associated with breast cancer that is proximal to an incident

polymorphic variation associated with breast cancer, which comprises identifying a polymorphic variant proximal to the incident polymorphic variant associated with breast cancer, where the incident polymorphic variant is in a nucleotide sequence set forth in SEQ ID NO: 1-4. The nucleotide sequence often comprises a polynucleotide sequence selected from the group consisting of (a) a nucleotide sequence set forth in SEQ ID NO: 1-4; (b) a nucleotide sequence which encodes a polypeptide having an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4 or a nucleotide sequence about 90% or more identical to the nucleotide sequence set forth in SEQ ID NO: 1-4; and (d) a fragment of a nucleotide sequence of (a), (b), or (c), often a fragment that includes a polymorphic site associated with breast cancer. The presence or absence of an association of the proximal polymorphic variant with breast cancer then is determined using a known association method, such as a method described in the Examples hereafter. In an embodiment, the incident polymorphic variant is described in SEQ ID NO: 1-4. In another embodiment, the proximal polymorphic variant identified sometimes is a publicly disclosed polymorphic variant, which for example, sometimes is published in a publicly available database. In other embodiments, the polymorphic variant identified is not publicly disclosed and is discovered using a known method, including, but not limited to, sequencing a region surrounding the incident polymorphic variant in a group of nucleic acid samples. Thus, multiple polymorphic variants proximal to an incident polymorphic variant are associated with breast cancer using this method.

[0048] The proximal polymorphic variant often is identified in a region surrounding the incident polymorphic variant. In certain embodiments, this surrounding region is about 50 kb flanking the first polymorphic variant (*e.g.* about 50 kb 5' of the first polymorphic variant and about 50 kb 3' of the first polymorphic variant), and the region sometimes is composed of shorter flanking sequences, such as flanking sequences of about 40 kb, about 30 kb, about 25 kb, about 20 kb, about 15 kb, about 10 kb, about 7 kb, about 5 kb, or about 2 kb 5' and 3' of the incident polymorphic variant. In other embodiments, the region is composed of longer flanking sequences, such as flanking sequences of about 55 kb, about 60 kb, about 65 kb, about 70 kb, about 75 kb, about 80 kb, about 85 kb, about 90 kb, about 95 kb, or about 100 kb 5' and 3' of the incident polymorphic variant.

[0049] In certain embodiments, polymorphic variants associated with breast cancer are identified iteratively. For example, a first proximal polymorphic variant is associated with breast cancer using the methods described above and then another polymorphic variant proximal to the first proximal polymorphic variant is identified (*e.g.*, publicly disclosed or discovered) and the presence or absence of an association of one or more other polymorphic variants proximal to the first proximal polymorphic variant with breast cancer is determined.

[0050] The methods described herein are useful for identifying or discovering additional polymorphic variants that may be used to further characterize a gene, region or loci associated with a

condition, a disease (*e.g.*, breast cancer), or a disorder. For example, allelotyping or genotyping data from the additional polymorphic variants may be used to identify a functional mutation or a region of linkage disequilibrium.

[0051] In certain embodiments, polymorphic variants identified or discovered within a region comprising the first polymorphic variant associated with breast cancer are genotyped using the genetic methods and sample selection techniques described herein, and it can be determined whether those polymorphic variants are in linkage disequilibrium with the first polymorphic variant. The size of the region in linkage disequilibrium with the first polymorphic variant also can be assessed using these genotyping methods. Thus, provided herein are methods for determining whether a polymorphic variant is in linkage disequilibrium with a first polymorphic variant associated with breast cancer, and such information can be used in prognosis methods described herein.

Isolated *DLG1, KIAA0783, DPF3* or *CENPC1* Nucleic Acids

[0052] Featured herein are isolated *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acids, which include the nucleic acid having the nucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11, nucleic acid variants, and substantially identical nucleic acids of the foregoing. Nucleotide sequences of the *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acids sometimes are referred to herein as "*DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequences." A "*DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid variant" refers to one allele that may have one or more different polymorphic variations as compared to another allele in another subject or the same subject. A polymorphic variation in the *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid variant may be represented on one or both strands in a double-stranded nucleic acid or on one chromosomal complement (heterozygous) or both chromosomal complements (homozygous).

[0053] As used herein, the term "nucleic acid" includes DNA molecules (e.g., a complementary DNA (cDNA) and genomic DNA (gDNA)) and RNA molecules (e.g., mRNA, rRNA, and tRNA) and analogs of DNA or RNA, for example, by use of nucleotide analogs. The nucleic acid molecule can be single-stranded and it is often double-stranded. The term "isolated or purified nucleic acid" refers to nucleic acids that are separated from other nucleic acids present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term "isolated" includes nucleic acids which are separated from the chromosome with which the genomic DNA is naturally associated. An "isolated" nucleic acid is often free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant

techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term "*DLG1, KIAA0783, DPF3* or *CENPC1* gene" refers to a nucleotide sequence that encodes a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide.

[0054] Also included herein are nucleic acid fragments. These fragments typically are a nucleotide sequence identical to a nucleotide sequence in SEQ ID NO: 1-8, a nucleotide sequence substantially identical to a nucleotide sequence in SEQ ID NO: 1-8, or a nucleotide sequence that is complementary to the foregoing. The nucleic acid fragment may be identical, substantially identical or homologous to a nucleotide sequence in an exon or an intron in SEQ ID NO: 1-4, and may encode a domain or part of a domain or motif of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide, sometimes the domains set forth in Figures 13-18. Sometimes, the fragment comprises the polymorphic variation described herein as being associated with breast cancer. The nucleic acid fragment sometimes is 50, 100, or 200 or fewer base pairs in length, and is sometimes about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3800, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 15000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 110000, 120000, 130000, 140000, 150000 or 160000 base pairs in length. A nucleic acid fragment complementary to a nucleotide sequence identical or substantially identical to the nucleotide sequence of SEQ ID NO: 1-8 and hybridizes to such a nucleotide sequence under stringent conditions often is referred to as a "probe." Nucleic acid fragments often include one or more polymorphic sites, or sometimes have an end that is adjacent to a polymorphic site as described hereafter.

[0055] An example of a nucleic acid fragment is an oligonucleotide. As used herein, the term "oligonucleotide" refers to a nucleic acid comprising about 8 to about 50 covalently linked nucleotides, often comprising from about 8 to about 35 nucleotides, and more often from about 10 to about 25 nucleotides. The backbone and nucleotides within an oligonucleotide may be the same as those of naturally occurring nucleic acids, or analogs or derivatives of naturally occurring nucleic acids, provided that oligonucleotides having such analogs or derivatives retain the ability to hybridize specifically to a nucleic acid comprising a targeted polymorphism. Oligonucleotides described herein may be used as hybridization probes or as components of prognostic or diagnostic assays, for example, as described herein.

[0056] Oligonucleotides are typically synthesized using standard methods and equipment, such as the ABI 3900 High Throughput DNA Synthesizer and the EXPEDITE™ 8909 Nucleic Acid Synthesizer, both of which are available from Applied Biosystems (Foster City, CA). Analogs and derivatives are exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684; 5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296; 6,140,482; WO 00/56746; WO 01/14398, and related publications. Methods for synthesizing oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the patent publications cited

above and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599; 5,962,674; 6,117,992; in WO 00/75372; and in related publications.

[0057] Oligonucleotides also may be linked to a second moiety. The second moiety may be an additional nucleotide sequence such as a tail sequence (e.g., a polyadenosine tail), an adapter sequence (e.g., phage M13 universal tail sequence), and others. Alternatively, the second moiety may be a non-nucleotide moiety such as a moiety which facilitates linkage to a solid support or a label to facilitate detection of the oligonucleotide. Such labels include, without limitation, a radioactive label, a fluorescent label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety may be attached to any position of the oligonucleotide, provided the oligonucleotide can hybridize to the nucleic acid comprising the polymorphism.

Uses for Nucleic Acid Sequences

[0058] Nucleic acid coding sequences depicted in SEQ ID NO: 1-8 may be used for diagnostic purposes for detection and control of polypeptide expression. Also, included herein are oligonucleotide sequences such as antisense RNA, small-interfering RNA (siRNA) and DNA molecules and ribozymes that function to inhibit translation of a polypeptide. Antisense techniques and RNA interference techniques are known in the art and are described herein.

[0059] Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by a endonucleolytic cleavage. Ribozymes may be engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences corresponding to or complementary to the nucleotide sequences set forth in SEQ ID NO: 1-8. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between fifteen (15) and twenty (20) ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

[0060] Antisense RNA and DNA molecules, siRNA and ribozymes may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs

that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

[0061] DNA encoding a polypeptide also may have a number of uses for the diagnosis of diseases, including breast cancer, resulting from aberrant expression of a target gene described herein. For example, the nucleic acid sequence may be used in hybridization assays of biopsies or autopsies to diagnose abnormalities of expression or function (*e.g.*, Southern or Northern blot analysis, in situ hybridization assays).

[0062] In addition, the expression of a polypeptide during embryonic development may also be determined using nucleic acid encoding the polypeptide. As addressed, *infra*, production of functionally impaired polypeptide can be the cause of various disease states, such as breast cancer. *In situ* hybridizations using polynucleotide probes may be employed to predict problems related to breast cancer. Further, as indicated, *infra*, administration of human active polypeptide, recombinantly produced as described herein, may be used to treat disease states related to functionally impaired polypeptide. Alternatively, gene therapy approaches may be employed to remedy deficiencies of functional polypeptide or to replace or compete with dysfunctional polypeptide.

Expression Vectors, Host Cells, and Genetically Engineered Cells

[0063] Provided herein are nucleic acid vectors, often expression vectors, which contain a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid, or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors may include replication defective retroviruses, adenoviruses and adeno-associated viruses for example.

[0064] A vector can include a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid in a form suitable for expression of the nucleic acid in a host cell. The recombinant expression vector typically includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. Expression vectors can be introduced into host cells to produce *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides, including fusion polypeptides, encoded by *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acids.

[0065] Recombinant expression vectors can be designed for expression of *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides in prokaryotic or eukaryotic cells. For example, *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides can be expressed in E. coli, insect cells (e.g., using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in

Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0066] Expression of polypeptides in prokaryotes is most often carried out in E. coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant polypeptide; 2) to increase the solubility of the recombinant polypeptide; and 3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, Gene 67: 31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0067] Purified fusion polypeptides can be used in screening assays and to generate antibodies specific for *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides. In a therapeutic embodiment, fusion polypeptide expressed in a retroviral expression vector is used to infect bone marrow cells that are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

[0068] Expressing the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide is often used to maximize recombinant polypeptide expression (Gottesman, S., Gene Expression Technology: Methods in Enzymology, Academic Press, San Diego, California 185: 119-128 (1990)). Another strategy is to alter the nucleotide sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in E. coli (Wada et al., Nucleic Acids Res. 20: 2111-2118 (1992)). Such alteration of nucleotide sequences can be carried out by standard DNA synthesis techniques.

[0069] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. Recombinant mammalian expression vectors are often capable of directing expression of the nucleic acid in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include an albumin promoter (liver-specific; Pinkert et al., Genes Dev. 1: 268-277 (1987)), lymphoid-specific promoters (Calame & Eaton, Adv. Immunol. 43: 235-275 (1988)), promoters of T cell receptors (Winoto & Baltimore, EMBO J. 8: 729-733 (1989))

promoters of immunoglobulins (Banerji et al., Cell 33: 729-740 (1983); Queen & Baltimore, Cell 33: 741-748 (1983)), neuron-specific promoters (e.g., the neurofilament promoter; Byrne & Ruddle, Proc. Natl. Acad. Sci. USA 86: 5473-5477 (1989)), pancreas-specific promoters (Edlund et al., Science 230: 912-916 (1985)), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are sometimes utilized, for example, the murine hox promoters (Kessel & Gruss, Science 249: 374-379 (1990)) and the a-fetopolypeptide promoter (Campes & Tilghman, Genes Dev. 3: 537-546 (1989)).

[0070] A *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid may also be cloned into an expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid cloned in the antisense orientation can be chosen for directing constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. Antisense expression vectors can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al., Antisense RNA as a molecular tool for genetic analysis, Reviews - Trends in Genetics, Vol. 1(1) (1986).

[0071] Also provided herein are host cells that include a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid within a recombinant expression vector or *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid sequence fragments which allow it to homologously recombine into a specific site of the host cell genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide can be expressed in bacterial cells such as E. coli, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0072] Vectors can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, transduction/infection, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0073] A host cell provided herein can be used to produce (i.e., express) a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Accordingly, further provided are methods for producing a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide using the host cells described herein. In one embodiment, the method includes culturing host cells into which a recombinant expression vector encoding a

DLG1, KIAA0783, DPF3 or *CENPC1* polypeptide has been introduced in a suitable medium such that a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide is produced. In another embodiment, the method further includes isolating a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide from the medium or the host cell.

[0074] Also provided are cells or purified preparations of cells which include a *DLG1, KIAA0783, DPF3* or *CENPC1* transgene, or which otherwise misexpress *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Cell preparations can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In certain embodiments, the cell or cells include a *DLG1, KIAA0783, DPF3* or *CENPC1* transgene (e.g., a heterologous form of a *DLG1, KIAA0783, DPF3* or *CENPC1* such as a human gene expressed in non-human cells). The *DLG1, KIAA0783, DPF3* or *CENPC1* transgene can be misexpressed, e.g., overexpressed or underexpressed. In other embodiments, the cell or cells include a gene which misexpress an endogenous *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide (e.g., expression of a gene is disrupted, also known as a knockout). Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed *DLG1, KIAA0783, DPF3* or *CENPC1* alleles or for use in drug screening. Also provided are human cells (e.g., a hematopoietic stem cells) transformed with a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid.

[0075] Also provided are cells or a purified preparation thereof (e.g., human cells) in which an endogenous *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid is under the control of a regulatory sequence that does not normally control the expression of the endogenous *DLG1, KIAA0783, DPF3* or *CENPC1* gene. The expression characteristics of an endogenous gene within a cell (e.g., a cell line or microorganism) can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous *DLG1, KIAA0783, DPF3* or *CENPC1* gene. For example, an endogenous *DLG1, KIAA0783, DPF3* or *CENPC1* gene (e.g., a gene which is "transcriptionally silent," not normally expressed, or expressed only at very low levels) may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

Transgenic Animals

[0076] Non-human transgenic animals that express a heterologous *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide (e.g., expressed from a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid isolated from another organism) can be generated. Such animals are useful for studying the function and/or activity of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide and for identifying and/or evaluating modulators of *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid and *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide activity. As used herein, a "transgenic animal" is a non-human animal

such as a mammal (e.g., a non-human primate such as chimpanzee, baboon, or macaque; an ungulate such as an equine, bovine, or caprine; or a rodent such as a rat, a mouse, or an Israeli sand rat), a bird (e.g., a chicken or a turkey), an amphibian (e.g., a frog, salamander, or newt), or an insect (e.g., Drosophila melanogaster), in which one or more of the cells of the animal includes a *DLG1, KIAA0783, DPF3* or *CENPC1* transgene. A transgene is exogenous DNA or a rearrangement (e.g., a deletion of endogenous chromosomal DNA) that is often integrated into or occurs in the genome of cells in a transgenic animal. A transgene can direct expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, and other transgenes can reduce expression (e.g., a knockout). Thus, a transgenic animal can be one in which an endogenous *DLG1, KIAA0783, DPF3* or *CENPC1* gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (e.g., an embryonic cell of the animal) prior to development of the animal.

[0077] Intronic sequences and polyadenylation signals can also be included in the transgene to increase expression efficiency of the transgene. One or more tissue-specific regulatory sequences can be operably linked to a *DLG1, KIAA0783, DPF3* or *CENPC1* transgene to direct expression of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide to particular cells. A transgenic founder animal can be identified based upon the presence of a *DLG1, KIAA0783, DPF3* or *CENPC1* transgene in its genome and/or expression of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide can further be bred to other transgenic animals carrying other transgenes.

[0078] *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides can be expressed in transgenic animals or plants by introducing, for example, a nucleic acid encoding the polypeptide into the genome of an animal. In certain embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Also included is a population of cells from a transgenic animal.

DLG1, KIAA0783, DPF3 and *CENPC1* Polypeptides

[0079] Featured herein are isolated *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides, which include polypeptides having amino acid sequences set forth in SEQ ID NO: 9-12, and substantially identical polypeptides thereof. Such polypeptides sometimes are proteins or peptides. A *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide is a polypeptide encoded by a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid, where one nucleic acid can encode one or more different polypeptides. An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of a *DLG1, KIAA0783, DPF3* or

CENPC1 polypeptide or *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide variant having less than about 30%, 20%, 10% and sometimes 5% (by dry weight), of non-*DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide (also referred to herein as a "contaminating protein"), or of chemical precursors or non-*DLG1, KIAA0783, DPF3* or *CENPC1* chemicals. When the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or a biologically active portion thereof is recombinantly produced, it is also often substantially free of culture medium, specifically, where culture medium represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the polypeptide preparation. Isolated or purified *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide preparations are sometimes 0.01 milligrams or more or 0.1 milligrams or more, and often 1.0 milligrams or more and 10 milligrams or more in dry weight. In specific embodiments, a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide comprises a glutamine at amino acid position 278 in SEQ ID NO: 9 or a glycine at amino acid position 389 in SEQ ID NO: 12.

[0080] In another aspect, featured herein are *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides and biologically active or antigenic fragments thereof that are useful as reagents or targets in assays applicable to prevention, treatment or diagnosis of breast cancer. In another embodiment, provided herein are *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides having a *DLG1, KIAA0783, DPF3* or *CENPC1* activity or activities.

[0081] Further included herein are *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide fragments. The polypeptide fragment may be a domain or part of a domain of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. The polypeptide fragment is often 50 or fewer, 100 or fewer, or 200 or fewer amino acids in length, and is sometimes 300, 400, 500, 600, 700, or 900 or fewer amino acids in length. In certain embodiments, the polypeptide fragment comprises, consists essentially of, or consists of, at least 6 consecutive amino acids and not more than 1211 consecutive amino acids of SEQ ID NO: 9-12, or the polypeptide fragment comprises, consists essentially of, or consists of, at least 6 consecutive amino acids and not more than 543 consecutive amino acids of SEQ ID NO: 9-12.

[0082] *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides described herein can be used as immunogens to produce anti-*DLG1, KIAA0783, DPF3* or *CENPC1* antibodies in a subject, to purify *DLG1, KIAA0783, DPF3* or *CENPC1* ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of *DLG1, KIAA0783, DPF3* or *CENPC1* with a *DLG1, KIAA0783, DPF3* or *CENPC1* substrate. Full-length *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides and polynucleotides encoding the same may be specifically substituted for a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide fragment or polynucleotide encoding the same in any embodiment described herein.

[0083] Substantially identical polypeptides may depart from the amino acid sequences set forth in SEQ ID NO: 9-12 in different manners. For example, conservative amino acid modifications may be introduced at one or more positions in the amino acid sequences of SEQ ID NO: 9-12. A "conservative amino acid substitution" is one in which the amino acid is replaced by another amino

acid having a similar structure and/or chemical function. Families of amino acid residues having similar structures and functions are well known. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Also, essential and non-essential amino acids may be replaced. A "non-essential" amino acid is one that can be altered without abolishing or substantially altering the biological function of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide, whereas altering an "essential" amino acid abolishes or substantially alters the biological function of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Amino acids that are conserved among *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides are typically essential amino acids.

[0084] Also, *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides and polypeptide variants may exist as chimeric or fusion polypeptides. As used herein, a *DLG1, KIAA0783, DPF3* or *CENPC1* "chimeric polypeptide" or "fusion polypeptide" includes a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide linked to a non-*DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. A "non-*DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially identical to the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide, which includes, for example, a polypeptide that is different from the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide and derived from the same or a different organism. The *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide in the fusion polypeptide can correspond to an entire or nearly entire *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or a fragment thereof. The non-*DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide can be fused to the N-terminus or C-terminus of the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide.

[0085] Fusion polypeptides can include a moiety having high affinity for a ligand. For example, the fusion polypeptide can be a GST-*DLG1, KIAA0783, DPF3* or *CENPC1* fusion polypeptide in which the *DLG1, KIAA0783, DPF3* or *CENPC1* sequences are fused to the C-terminus of the GST sequences, or a polyhistidine-*DLG1, KIAA0783, DPF3* or *CENPC1* fusion polypeptide in which the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide is fused at the N- or C-terminus to a string of histidine residues. Such fusion polypeptides can facilitate purification of recombinant *DLG1, KIAA0783, DPF3* or *CENPC1*. Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide), and a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid can be cloned into an expression vector such that the fusion moiety is linked in-frame to the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Further, the fusion polypeptide can be a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression, secretion, cellular internalization, and cellular localization of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide can be

increased through use of a heterologous signal sequence. Fusion polypeptides can also include all or a part of a serum polypeptide (e.g., an IgG constant region or human serum albumin).

[0086] *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides or fragments thereof can be incorporated into pharmaceutical compositions and administered to a subject in vivo. Administration of these *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides can be used to affect the bioavailability of a *DLG1, KIAA0783, DPF3* or *CENPC1* substrate and may effectively increase or decrease *DLG1, KIAA0783, DPF3* or *CENPC1* biological activity in a cell or effectively supplement dysfunctional or hyperactive *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. *DLG1, KIAA0783, DPF3* or *CENPC1* fusion polypeptides may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide; (ii) mis-regulation of the *DLG1, KIAA0783, DPF3* or *CENPC1* gene; and (iii) aberrant post-translational modification of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Also, *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides can be used as immunogens to produce anti-*DLG1, KIAA0783, DPF3* or *CENPC1* antibodies in a subject, to purify *DLG1, KIAA0783, DPF3* or *CENPC1* ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of *DLG1, KIAA0783, DPF3* or *CENPC1* with a *DLG1, KIAA0783, DPF3* or *CENPC1* substrate. Preferably, said *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides are used in screening assays to identify molecules which inhibit the interaction of *DLG1, KIAA0783, DPF3* or *CENPC1*.

[0087] In addition, polypeptides can be chemically synthesized using techniques known in the art (See, *e.g.*, Creighton, 1983 Proteins. New York, N.Y.: W. H. Freeman and Company; and Hunkapiller *et al.*, (1984) Nature July 12 -18;310(5973):105-11). For example, a relative short polypeptide fragment can be synthesized by use of a peptide synthesizer. Furthermore, if desired, non-classical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the fragment sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, a-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, g-Abu, e-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, b-alanine, fluoroamino acids, designer amino acids such as b-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

[0088] Also included are polypeptide fragments which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, and the like. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin,

chymotrypsin, papain, V8 protease, $NaBH_4$; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; and the like.

[0089] Additional post-translational modifications include, for example, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. The polypeptide fragments may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the polypeptide.

[0090] Also provided are chemically modified polypeptide derivatives that may provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity. See U.S. Pat. No: 4,179,337. The chemical moieties for derivitization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

[0091] The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (*e.g.*, the duration of sustained release desired, the effects, if any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

[0092] The polyethylene glycol molecules (or other chemical moieties) should be attached to the polypeptide with consideration of effects on functional or antigenic domains of the polypeptide. There are a number of attachment methods available to those skilled in the art, *e.g.*, EP 0 401 384, herein incorporated by reference (coupling PEG to G-CSF), see also Malik *et al.* (1992) Exp Hematol. September;20(8):1028-35, reporting pegylation of GM-CSF using tresyl chloride). For example, polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as, a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues, glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. A polymer sometimes is attached at an amino group, such as attachment at the N-terminus or lysine group.

[0093] One may specifically desire proteins chemically modified at the N-terminus. Using polyethylene glycol as an illustration of the present composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, and the like), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (i.e., separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

Substantially Identical Nucleic Acids and Polypeptides

[0094] Nucleotide sequences and polypeptide sequences that are substantially identical to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence and the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide sequences encoded by those nucleotide sequences are included herein. The term "substantially identical" as used herein refers to two or more nucleic acids or polypeptides sharing one or more identical nucleotide sequences or polypeptide sequences, respectively. Included are nucleotide sequences or polypeptide sequences that are 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more (each often within a 1%, 2%, 3% or 4% variability) or more identical to the nucleotide sequences in SEQ ID NO: 1-8 or the encoded *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide amino acid sequences. One test for determining whether two nucleic acids are substantially identical is to determine the percent of identical nucleotide sequences or polypeptide sequences shared between the nucleic acids or polypeptides.

[0095] Calculations of sequence identity are often performed as follows. Sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences,

taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

[0096] Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, *CABIOS 4:* 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman & Wunsch, *J. Mol. Biol. 48:* 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the http address www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http address www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0097] Another manner for determining if two nucleic acids are substantially identical is to assess whether a polynucleotide homologous to one nucleic acid will hybridize to the other nucleic acid under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. , 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

[0098] An example of a substantially identical nucleotide sequence to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence is one that has a different nucleotide sequence but still encodes the same polypeptide sequence encoded by the *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence. Another example is a nucleotide sequence that encodes a polypeptide having a polypeptide

sequence that is more than 70% or more identical to, sometimes 75% or more, 80% or more, or 85% or more identical to, and often 90% or more and 95% or more identical to a polypeptide sequence encoded by a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence.

[0099] *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequences and *DLG1, KIAA0783, DPF3* or *CENPC1* amino acid sequences can be used as "query sequences" to perform a search against public databases to identify other family members or related sequences, for example. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al., J. Mol. Biol. 215:* 403-10 (1990). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleotide sequences from SEQ ID NO: 1-8. BLAST polypeptide searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to polypeptides encoded by a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al., Nucleic Acids Res. 25(17):* 3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used (*see* the http address www.ncbi.nlm.nih.gov).

[0100] A nucleic acid that is substantially identical to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence may include polymorphic sites at positions equivalent to those described herein when the sequences are aligned. For example, using the alignment procedures described herein, SNPs in a sequence substantially identical to a sequence in SEQ ID NO: 1-8 can be identified at nucleotide positions that match (*i.e.*, align) with nucleotides at SNP positions in the nucleotide sequence of SEQ ID NO: 1-8. Also, where a polymorphic variation results in an insertion or deletion, insertion or deletion of a nucleotide sequence from a reference sequence can change the relative positions of other polymorphic sites in the nucleotide sequence.

[0101] Substantially identical nucleotide and polypeptide sequences include those that are naturally occurring, such as allelic variants (same locus), splice variants, homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be generated by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product). Orthologs, homologs, allelic variants, and splice variants can be identified using methods known in the art. These variants normally comprise a nucleotide sequence encoding a polypeptide that is 50% or more, about 55% or more, often about 70-75% or more, more often about 80-85% or more, and typically about 90-95% or more identical to the amino acid sequences of target polypeptides or a fragment thereof. Such nucleic acid molecules readily can be identified as being able to hybridize under stringent conditions to a nucleotide sequence in SEQ ID NO: 1-8 or a

fragment thereof. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of a nucleotide sequence in SEQ ID NO: 1-8 can be identified by mapping the sequence to the same chromosome or locus as the nucleotide sequence in SEQ ID NO: 1-8.

[0102] Also, substantially identical nucleotide sequences may include codons that are altered with respect to the naturally occurring sequence for enhancing expression of a target polypeptide in a particular expression system. For example, the nucleic acid can be one in which one or more codons are altered, and often 10% or more or 20% or more of the codons are altered for optimized expression in bacteria (*e.g.*, *E. coli.*), yeast (*e.g.*, *S. cervesiae*), human (*e.g.*, 293 cells), insect, or rodent (*e.g.*, hamster) cells.

Methods for Identifying Subjects at Risk of Breast Cancer and Breast Cancer Risk in a Subject

[0103] Methods for prognosing and diagnosing breast cancer in subjects are provided herein. These methods include detecting the presence or absence of one or more polymorphic variations associated with breast cancer in a nucleotide sequence set forth in SEQ ID NO: 1-4, or substantially identical sequence thereof, in a sample from a subject, where the presence of a polymorphic variant is indicative of a risk of breast cancer.

[0104] Thus, featured herein is a method for detecting a subject at risk of breast cancer or the risk of breast cancer in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with breast cancer at a polymorphic site in a nucleotide sequence set forth in SEQ ID NO: 1-4 in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence set forth in SEQ ID NO: 1-4; (b) a nucleotide sequence which encodes a polypeptide having an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4 or a nucleotide sequence about 90% or more identical to the nucleotide sequence set forth in SEQ ID NO: 1-4; and (d) a fragment of a nucleotide sequence of (a), (b), or (c), often a fragment that includes a polymorphic site associated with breast cancer; whereby the presence of the polymorphic variation is indicative of a risk of breast cancer in the subject. In certain embodiments, determining the presence of a combination of two or more polymorphic variants associated with breast cancer in one or more nucleotide sequences of the sample is determined to identify a subject at risk of breast cancer and/or risk of breast cancer.

[0105] A risk of developing aggressive forms of breast cancer likely to metastasize or invade surrounding tissues (e.g., Stage IIIA, IIIB, and IV breast cancers), and subjects at risk of developing aggressive forms of breast cancer also may be identified by the methods described herein. These methods include collecting phenotype information from subjects having breast cancer, which includes the stage of progression of the breast cancer, and performing a secondary phenotype analysis to detect

the presence or absence of one or more polymorphic variations associated with a particular stage form of breast cancer. Thus, detecting the presence or absence of one or more polymorphic variations in a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence associated with a late stage form of breast cancer often is prognostic and/or diagnostic of an aggressive form of the cancer.

[0106] Results from prognostic tests may be combined with other test results to diagnose breast cancer. For example, prognostic results may be gathered, a patient sample may be ordered based on a determined predisposition to breast cancer, the patient sample is analyzed, and the results of the analysis may be utilized to diagnose breast cancer. Also breast cancer diagnostic methods can be developed from studies used to generate prognostic/diagnostic methods in which populations are stratified into subpopulations having different progressions of breast cancer. In another embodiment, prognostic results may be gathered; a patient's risk factors for developing breast cancer analyzed (*e.g.*, age, race, family history, age of first menstrual cycle, age at birth of first child); and a patient sample may be ordered based on a determined predisposition to breast cancer. In an alternative embodiment, the results from predisposition analyses described herein may be combined with other test results indicative of breast cancer, which were previously, concurrently, or subsequently gathered with respect to the predisposition testing. In these embodiments, the combination of the prognostic test results with other test results can be probative of breast cancer, and the combination can be utilized as a breast cancer diagnostic. The results of any test indicative of breast cancer known in the art may be combined with the methods described herein. Examples of such tests are mammography (*e.g.*, a more frequent and/or earlier mammography regimen may be prescribed); breast biopsy and optionally a biopsy from another tissue; breast ultrasound and optionally an ultrasound analysis of another tissue; breast magnetic resonance imaging (MRI) and optionally an MRI analysis of another tissue; electrical impedance (T-scan) analysis of breast and optionally of another tissue; ductal lavage; nuclear medicine analysis (*e.g.*, scintimammography); *BRCA1* and/or *BRCA2* sequence analysis results; and thermal imaging of the breast and optionally of another tissue. Testing may be performed on tissue other than breast to diagnose the occurrence of metastasis (*e.g.*, testing of the lymph node).

[0107] Risk of breast cancer sometimes is expressed as a probability, such as an odds ratio, percentage, or risk factor. The risk is based upon the presence or absence of one or more polymorphic variants described herein, and also may be based in part upon phenotypic traits of the individual being tested. Methods for calculating predispositions based upon patient data are well known (*see, e.g.*, Agresti, *Categorical Data Analysis*, 2nd Ed. 2002. Wiley). Allelotyping and genotyping analyses may be carried out in populations other than those exemplified herein to enhance the predictive power of the prognostic method. These further analyses are executed in view of the exemplified procedures described herein, and may be based upon the same polymorphic variations or additional polymorphic variations. Risk determinations for breast cancer are useful in a variety of applications. In one embodiment, breast cancer risk determinations are used by clinicians to direct appropriate detection, preventative and treatment procedures to subjects who most require these. In another embodiment,

breast cancer risk determinations are used by health insurers for preparing actuarial tables and for calculating insurance premiums.

[0108] The nucleic acid sample typically is isolated from a biological sample obtained from a subject. For example, nucleic acid can be isolated from blood, saliva, sputum, urine, cell scrapings, and biopsy tissue. The nucleic acid sample can be isolated from a biological sample using standard techniques, such as the technique described in Example 2. As used herein, the term "subject" refers primarily to humans but also refers to other mammals such as dogs, cats, and ungulates (*e.g.*, cattle, sheep, and swine). Subjects also include avians (*e.g.*, chickens and turkeys), reptiles, and fish (*e.g.*, salmon), as embodiments described herein can be adapted to nucleic acid samples isolated from any of these organisms. The nucleic acid sample may be isolated from the subject and then directly utilized in a method for determining the presence of a polymorphic variant, or alternatively, the sample may be isolated and then stored (*e.g.*, frozen) for a period of time before being subjected to analysis.

[0109] The presence or absence of a polymorphic variant is determined using one or both chromosomal complements represented in the nucleic acid sample. Determining the presence or absence of a polymorphic variant in both chromosomal complements represented in a nucleic acid sample from a subject having a copy of each chromosome is useful for determining the zygosity of an individual for the polymorphic variant (*i.e.*, whether the individual is homozygous or heterozygous for the polymorphic variant). Any oligonucleotide-based diagnostic may be utilized to determine whether a sample includes the presence or absence of a polymorphic variant in a sample. For example, primer extension methods, ligase sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO 01/27326), mismatch sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), microarray sequence determination methods, restriction fragment length polymorphism (RFLP), single strand conformation polymorphism detection (SSCP) (*e.g.*, U.S. Pat. Nos. 5,891,625 and 6,013,499), PCR-based assays (*e.g.*, TAQMAN® PCR System (Applied Biosystems)), and nucleotide sequencing methods may be used.

[0110] Oligonucleotide extension methods typically involve providing a pair of oligonucleotide primers in a polymerase chain reaction (PCR) or in other nucleic acid amplification methods for the purpose of amplifying a region from the nucleic acid sample that comprises the polymorphic variation. One oligonucleotide primer is complementary to a region 3' of the polymorphism and the other is complementary to a region 5' of the polymorphism. A PCR primer pair may be used in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202, 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GENEAMP® Systems available from Applied Biosystems. Also, those of ordinary skill in the art will be able to design oligonucleotide primers based upon a nucleotide sequence set forth in SEQ ID NO: 1-4 without undue experimentation using knowledge readily available in the art.

[0111] Also provided is an extension oligonucleotide that hybridizes to the amplified fragment adjacent to the polymorphic variation. As used herein, the term "adjacent" refers to the 3' end of the extension oligonucleotide being often 1 nucleotide from the 5' end of the polymorphic site, and sometimes 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine whether the polymorphic variant is present. Oligonucleotide extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; and 6,194,144, and a method often utilized is described herein in Example 2. Multiple extension oligonucleotides may be utilized in one reaction, which is referred to herein as "multiplexing."

[0112] A microarray can be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A microarray may include any oligonucleotides described herein, and methods for making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; and WO 01/29259. The microarray typically comprises a solid support and the oligonucleotides may be linked to this solid support by covalent bonds or by non-covalent interactions. The oligonucleotides may also be linked to the solid support directly or by a spacer molecule. A microarray may comprise one or more oligonucleotides complementary to a polymorphic site set forth in SEQ ID NO: 1-4 or below.

[0113] A kit also may be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A kit often comprises one or more pairs of oligonucleotide primers useful for amplifying a fragment of a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence or a substantially identical sequence thereof, where the fragment includes a polymorphic site. The kit sometimes comprises a polymerizing agent, for example, a thermostable nucleic acid polymerase such as one disclosed in U.S. Pat. Nos. 4,889,818 or 6,077,664. Also, the kit often comprises an elongation oligonucleotide that hybridizes to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence in a nucleic acid sample adjacent to the polymorphic site. Where the kit includes an elongation oligonucleotide, it also often comprises chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, provided that such analogs are substrates for a thermostable nucleic acid polymerase and can be incorporated into a nucleic acid chain elongated from the extension oligonucleotide. Along with chain elongating nucleotides would be one or more chain terminating nucleotides such as ddATP, ddTTP, ddGTP,

ddCTP, and the like. In an embodiment, the kit comprises one or more oligonucleotide primer pairs, a polymerizing agent, chain elongating nucleotides, at least one elongation oligonucleotide, and one or more chain terminating nucleotides. Kits optionally include buffers, vials, microtiter plates, and instructions for use.

[0114] An individual identified as being at risk of breast cancer may be heterozygous or homozygous with respect to the allele associated with a higher risk of breast cancer. A subject homozygous for an allele associated with an increased risk of breast cancer is at a comparatively high risk of breast cancer, a subject heterozygous for an allele associated with an increased risk of breast cancer is at a comparatively intermediate risk of breast cancer, and a subject homozygous for an allele associated with a decreased risk of breast cancer is at a comparatively low risk of breast cancer. A genotype may be assessed for a complementary strand, such that the complementary nucleotide at a particular position is detected.

[0115] Also featured are methods for determining risk of breast cancer and/or identifying a subject at risk of breast cancer by contacting a polypeptide or protein encoded by a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence from a subject with an antibody that specifically binds to an epitope associated with increased risk of breast cancer in the polypeptide. In certain embodiments, the antibody specifically binds to an epitope that comprises a glutamine at amino acid position 278 in SEQ ID NO: 9 or a glycine at amino acid position 389 in SEQ ID NO: 12.

Applications of Prognostic and Diagnostic Results to Pharmacogenomic Methods

[0116] Pharmacogenomics is a discipline that involves tailoring a treatment for a subject according to the subject's genotype. For example, based upon the outcome of a prognostic test described herein, a clinician or physician may target pertinent information and preventative or therapeutic treatments to a subject who would be benefited by the information or treatment and avoid directing such information and treatments to a subject who would not be benefited (*e.g.*, the treatment has no therapeutic effect and/or the subject experiences adverse side effects). As therapeutic approaches for breast cancer continue to evolve and improve, the goal of treatments for breast cancer related disorders is to intervene even before clinical signs (*e.g.*, identification of lump in the breast) first manifest. Thus, genetic markers associated with susceptibility to breast cancer prove useful for early diagnosis, prevention and treatment of breast cancer.

[0117] The following is an example of a pharmacogenomic embodiment. A particular treatment regimen can exert a differential effect depending upon the subject's genotype. Where a candidate therapeutic exhibits a significant interaction with a major allele and a comparatively weak interaction with a minor allele (*e.g.*, an order of magnitude or greater difference in the interaction), such a therapeutic typically would not be administered to a subject genotyped as being homozygous for the minor allele, and sometimes not administered to a subject genotyped as being heterozygous for the minor allele. In another example, where a candidate therapeutic is not significantly toxic when

administered to subjects who are homozygous for a major allele but is comparatively toxic when administered to subjects heterozygous or homozygous for a minor allele, the candidate therapeutic is not typically administered to subjects who are genotyped as being heterozygous or homozygous with respect to the minor allele.

[0118] The methods described herein are applicable to pharmacogenomic methods for detecting, preventing, alleviating and/or treating breast cancer. For example, a nucleic acid sample from an individual may be subjected to a genetic test described herein. Where one or more polymorphic variations associated with increased risk of breast cancer are identified in a subject, information for detecting, preventing or treating breast cancer and/or one or more breast cancer detection, prevention and/or treatment regimens then may be directed to and/or prescribed to that subject.

[0119] In certain embodiments, a detection, prevenative and/or treatment regimen is specifically prescribed and/or administered to individuals who will most benefit from it based upon their risk of developing breast cancer assessed by the methods described herein. Thus, provided are methods for identifying a subject at risk of breast cancer and then prescribing a detection, therapeutic or preventative regimen to individuals identified as being at risk of breast cancer. Thus, certain embodiments are directed to methods for treating breast cancer in a subject, reducing risk of breast cancer in a subject, or early detection of breast cancer in a subject, which comprise: detecting the presence or absence of a polymorphic variant associated with breast cancer in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence set forth in SEQ ID NO: 1-4; (b) a nucleotide sequence which encodes a polypeptide having an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4 or a nucleotide sequence about 90% or more identical to the nucleotide sequence set forth in SEQ ID NO: 1-4; and (d) a fragment of a nucleotide sequence of (a), (b), or (c), sometimes comprising a polymorphic site associated with breast cancer; and prescribing or administering a breast cancer treatment regimen, preventative regimen and/or detection regimen to a subject from whom the sample originated where the presence of one or more polymorphic variations associated with breast cancer are detected in the nucleotide sequence. In these methods, genetic results may be utilized in combination with other test results to diagnose breast cancer as described above. Other test results include but are not limited to mammography results, imaging results, biopsy results and results from *BRCA1* or *BRAC2* test results, as described above.

[0120] Detection regimens include one or more mammography procedures, a regular mammography regimen (*e.g.*, once a year, or once every six, four, three or two months); an early mammography regimen (*e.g.*, mammography tests are performed beginning at age 25, 30, or 35); one or more biopsy procedures (*e.g.*, a regular biopsy regimen beginning at age 40); breast biopsy and biopsy from other tissue; breast ultrasound and optionally ultrasound analysis of another tissue; breast

magnetic resonance imaging (MRI) and optionally MRI analysis of another tissue; electrical impedance (T-scan) analysis of breast and optionally another tissue; ductal lavage; nuclear medicine analysis (*e.g.*, scintimammography); *BRCA1* and/or *BRCA2* sequence analysis results; and/or thermal imaging of the breast and optionally another tissue.

[0121] Treatments sometimes are preventative (*e.g.*, is prescribed or administered to reduce the probability that a breast cancer associated condition arises or progresses), sometimes are therapeutic, and sometimes delay, alleviate or halt the progression of breast cancer. Any known preventative or therapeutic treatment for alleviating or preventing the occurrence of breast cancer is prescribed and/or administered. For example, certain preventative treatments often are prescribed to subjects having a predisposition to breast cancer and where the subject is not diagnosed with breast cancer or is diagnosed as having symptoms indicative of early stage breast cancer (*e.g.*, stage I). For subjects not diagnosed as having breast cancer, any preventative treatments known in the art can be prescribed and administered, which include selective hormone receptor modulators (*e.g.*, selective estrogen receptor modulators (SERMs) such as tamoxifen, reloxifene, and toremifene); compositions that prevent production of hormones (*e.g.*, aramotase inhibitors that prevent the production of estrogen in the adrenal gland, such as exemestane, letrozole, anastrozol, groserelin, and megestrol); other hormonal treatments (*e.g.*, goserelin acetate and fulvestrant); biologic response modifiers such as antibodies (*e.g.*, trastuzumab (herceptin/HER2)); surgery (*e.g.*, lumpectomy and mastectomy); drugs that delay or halt metastasis (*e.g.*, pamidronate disodium); and alternative/complementary medicine (*e.g.*, acupuncture, acupressure, moxibustion, qi gong, reiki, ayurveda, vitamins, minerals, and herbs (*e.g.*, astragalus root, burdock root, garlic, green tea, and licorice root)).

[0122] The use of breast cancer treatments are well known in the art, and include surgery, chemotherapy and/or radiation therapy. Any of the treatments may be used in combination to treat or prevent breast cancer (*e.g.*, surgery followed by radiation therapy or chemotherapy). Examples of chemotherapy combinations used to treat breast cancer include: cyclophosphamide (Cytoxan), methotrexate (Amethopterin, Mexate, Folex), and fluorouracil (Fluorouracil, 5-Fu, Adrucil), which is referred to as CMF; cyclophosphamide, doxorubicin (Adriamycin), and fluorouracil, which is referred to as CAF; and doxorubicin (Adriamycin) and cyclophosphamide, which is referred to as AC.

[0123] As breast cancer preventative and treatment information can be specifically targeted to subjects in need thereof (*e.g.*, those at risk of developing breast cancer or those that have early signs of breast cancer), provided herein is a method for preventing or reducing the risk of developing breast cancer in a subject, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with breast cancer at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying a subject with a predisposition to breast cancer, whereby the presence of the polymorphic variation is indicative of a predisposition to breast cancer in the subject; and (c) if such a predisposition is identified, providing the subject with information about methods or products to prevent or reduce breast cancer or to delay the onset of breast cancer. Also provided is a method of

targeting information or advertising to a subpopulation of a human population based on the subpopulation being genetically predisposed to a disease or condition, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with breast cancer at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying the subpopulation of subjects in which the polymorphic variation is associated with breast cancer; and (c) providing information only to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition.

[0124] Pharmacogenomics methods also may be used to analyze and predict a response to a breast cancer treatment or a drug. For example, if pharmacogenomics analysis indicates a likelihood that an individual will respond positively to a breast cancer treatment with a particular drug, the drug may be administered to the individual. Conversely, if the analysis indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects. The response to a therapeutic treatment can be predicted in a background study in which subjects in any of the following populations are genotyped: a population that responds favorably to a treatment regimen, a population that does not respond significantly to a treatment regimen, and a population that responds adversely to a treatment regiment (*e.g.*, exhibits one or more side effects). These populations are provided as examples and other populations and subpopulations may be analyzed. Based upon the results of these analyses, a subject is genotyped to predict whether he or she will respond favorably to a treatment regimen, not respond significantly to a treatment regimen, or respond adversely to a treatment regimen.

[0125] The methods described herein also are applicable to clinical drug trials. One or more polymorphic variants indicative of response to an agent for treating breast cancer or to side effects to an agent for treating breast cancer may be identified using the methods described herein. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems. In certain embodiments, the agent for treating breast cancer described herein targets *DLG1, KIAA0783, DPF3* or *CENPC1* or a target in the *DLG1, KIAA0783, DPF3* or *CENPC1* pathway.

[0126] Thus, another embodiment is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: (a) obtaining a nucleic acid sample from an individual; (b) determining the identity of a polymorphic variation which is associated with a positive response to the treatment or the drug, or at least one polymorphic variation which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and (c) including the individual in the clinical trial if the nucleic acid sample contains said polymorphic

variation associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said polymorphic variation associated with a negative response to the treatment or the drug. In addition, the methods for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. The polymorphic variation may be in a sequence selected individually or in any combination from the group consisting of (i) a polynucleotide sequence set forth in SEQ ID NO: 1-4; (ii) a polynucleotide sequence that is 90% or more identical to a nucleotide sequence set forth in SEQ ID NO: 1-4; (iii) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence identical to or 90% or more identical to an amino acid sequence encoded by a nucleotide sequence set forth in SEQ ID NO: 1-4; and (iv) a fragment of a polynucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site. The including step (c) optionally comprises administering the drug or the treatment to the individual if the nucleic acid sample contains the polymorphic variation associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

[0127] Also provided herein is a method of partnering between a diagnostic/prognostic testing provider and a provider of a consumable product, which comprises: (a) the diagnostic/prognostic testing provider detects the presence or absence of a polymorphic variation associated with breast cancer at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) the diagnostic/prognostic testing provider identifies the subpopulation of subjects in which the polymorphic variation is associated with breast cancer; (c) the diagnostic/prognostic testing provider forwards information to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition; and (d) the provider of a consumable product forwards to the diagnostic test provider a fee every time the diagnostic/prognostic test provider forwards information to the subject as set forth in step (c) above.

Compositions Comprising Breast Cancer-Directed Molecules

[0128] Featured herein is a composition comprising a breast cancer cell and one or more molecules specifically directed and targeted to a nucleic acid comprising a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence or a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Such directed molecules include, but are not limited to, a compound that binds to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid or a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide; a RNAi or siRNA molecule having a strand complementary to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence; an antisense nucleic acid complementary to an RNA encoded by a *DLG1, KIAA0783, DPF3* or *CENPC1* DNA sequence; a ribozyme that hybridizes to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence; a nucleic acid aptamer that specifically binds a *DLG1,*

KIAA0783, DPF3 or *CENPC1* polypeptide; and an antibody that specifically binds to a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or binds to a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid. In certain embodiments, the antibody specifically binds to an epitope that comprises a glutamine at amino acid position 278 in SEQ ID NO: 9 or a glycine at amino acid position 389 in SEQ ID NO: 12. In specific embodiments, the breast cancer directed molecule interacts with a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid or polypeptide variant associated with breast cancer. In other embodiments, the breast cancer directed molecule interacts with a polypeptide involved in the *DLG1, KIAA0783, DPF3* or *CENPC1* signal pathway, or a nucleic acid encoding such a polypeptide. Polypeptides involved in the *DLG1, KIAA0783, DPF3* or *CENPC1* signal pathway are discussed herein.

[0129] Compositions sometimes include an adjuvant known to stimulate an immune response, and in certain embodiments, an adjuvant that stimulates a T-cell lymphocyte response. Adjuvants are known, including but not limited to an aluminum adjuvant (*e.g.*, aluminum hydroxide); a cytokine adjuvant or adjuvant that stimulates a cytokine response (*e.g.*, interleukin (IL)-12 and/or ?-interferon cytokines); a Freund-type mineral oil adjuvant emulsion (*e.g.*, Freund's complete or incomplete adjuvant); a synthetic lipoid compound; a copolymer adjuvant (*e.g.*, TitreMax); a saponin; Quil A; a liposome; an oil-in-water emulsion (*e.g.*, an emulsion stabilized by Tween 80 and pluronic polyoxyethlene/polyoxypropylene block copolymer (Syntex Adjuvant Formulation); TitreMax; detoxified endotoxin (MPL) and mycobacterial cell wall components (TDW, CWS) in 2% squalene (Ribi Adjuvant System)); a muramyl dipeptide; an immune-stimulating complex (ISCOM, *e.g.*, an Ag-modified saponin/cholesterol micelle that forms stable cage-like structure); an aqueous phase adjuvant that does not have a depot effect (*e.g.*, Gerbu adjuvant); a carbohydrate polymer (*e.g.*, AdjuPrime); L-tyrosine; a manide-oleate compound (*e.g.*, Montanide); an ethylene-vinyl acetate copolymer (*e.g.*, Elvax 40W1,2); or lipid A, for example. Such compositions are useful for generating an immune response against a breast cancer directed molecule (*e.g.*, an HLA-binding subsequence within a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4). In such methods, a peptide having an amino acid subsequence of a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4 is delivered to a subject, where the subsequence binds to an HLA molecule and induces a CTL lymphocyte response. The peptide sometimes is delivered to the subject as an isolated peptide or as a minigene in a plasmid that encodes the peptide. Methods for identifying HLA-binding subsequences in such polypeptides are known (*see e.g.*, publication WO02/20616 and PCT application US98/01373 for methods of identifying such sequences).

[0130] The breast cancer cell may be in a group of breast cancer cells and/or other types of cells cultured *in vitro* or in a tissue having breast cancer cells (*e.g.*, a melanocytic lesion) maintained *in vitro* or present in an animal *in vivo* (*e.g.*, a rat, mouse, ape or human). In certain embodiments, a composition comprises a component from a breast cancer cell or from a subject having a breast cancer cell instead of the breast cancer cell or in addition to the breast cancer cell, where the component

sometimes is a nucleic acid molecule (*e.g.*, genomic DNA), a protein mixture or isolated protein, for example. The aforementioned compositions have utility in diagnostic, prognostic and pharmacogenomic methods described previously and in breast cancer therapeutics described hereafter. Certain breast cancer molecules are described in greater detail below.

Compounds

[0131] Compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive (see, *e.g.*, Zuckermann *et al.*, J. Med. Chem.37: 2678-85 (1994)); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; "one-bead one-compound" library methods; and synthetic library methods using affinity chromatography selection. Biological library and peptoid library approaches are typically limited to peptide libraries, while the other approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, Anticancer Drug Des. 12: 145, (1997)). Examples of methods for synthesizing molecular libraries are described, for example, in DeWitt *et al.*, Proc. Natl. Acad. Sci. U.S.A. 90: 6909 (1993); Erb *et al.*, Proc. Natl. Acad. Sci. USA 91: 11422 (1994); Zuckermann *et al.*, J. Med. Chem. 37: 2678 (1994); Cho *et al.*, Science 261: 1303 (1993); Carrell *et al.*, Angew. Chem. Int. Ed. Engl. 33: 2059 (1994); Carell *et al.*, Angew. Chem. Int. Ed. Engl. 33: 2061 (1994); and in Gallop *et al.*, J. Med. Chem. 37: 1233 (1994).

[0132] Libraries of compounds may be presented in solution (*e.g.*, Houghten, Biotechniques 13: 412-421 (1992)), or on beads (Lam, Nature 354: 82-84 (1991)), chips (Fodor, Nature 364: 555-556 (1993)), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull *et al.*, Proc. Natl. Acad. Sci. USA 89: 1865-1869 (1992)) or on phage (Scott and Smith, Science 249: 386-390 (1990); Devlin, Science 249: 404-406 (1990); Cwirla *et al.*, Proc. Natl. Acad. Sci. 87: 6378-6382 (1990); Felici, J. Mol. Biol. 222: 301-310 (1991); Ladner supra.).

[0133] A compound sometimes alters expression and sometimes alters activity of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide and may be a small molecule. Small molecules include, but are not limited to, peptides, peptidomimetics (*e.g.*, peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Antisense Nucleic Acid Molecules, Ribozymes, RNAi, siRNA and Modified Nucleic Acid Molecules

[0134] An "antisense" nucleic acid refers to a nucleotide sequence complementary to a "sense" nucleic acid encoding a polypeptide, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire coding strand in SEQ ID NO: 1-8, or to a portion thereof or a substantially identical sequence thereof. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence in SEQ ID NO: 1-8 (*e.g.*, 5' and 3' untranslated regions).

[0135] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of an mRNA encoded by a nucleotide sequence in SEQ ID NO: 1-4 (e.g., SEQ ID NO: 6-11), and often the antisense nucleic acid is an oligonucleotide antisense to only a portion of a coding or noncoding region of the mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of the mRNA, *e.g.*, between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. The antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target a nucleotide sequence in SEQ ID NO: 1-8, often a variant associated with breast cancer, or a substantially identical sequence thereof. Among the variants, minor alleles and major alleles can be targeted, and those associated with a higher risk of breast cancer are often designed, tested, and administered to subjects.

[0136] An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using standard procedures. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0137] When utilized as therapeutics, antisense nucleic acids typically are administered to a subject (*e.g.*, by direct injection at a tissue site) or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a polypeptide and thereby inhibit expression of the polypeptide, for example, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then are administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to

receptors or antigens expressed on a selected cell surface, for example, by linking antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. Antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. Sufficient intracellular concentrations of antisense molecules are achieved by incorporating a strong promoter, such as a pol II or pol III promoter, in the vector construct.

[0138] Antisense nucleic acid molecules sometimes are *-anomeric nucleic acid molecules. An *-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual *-units, the strands run parallel to each other (Gaultier *et al.*, Nucleic Acids. Res. 15: 6625-6641 (1987)). Antisense nucleic acid molecules can also comprise a 2'-o-methylribonucleotide (Inoue *et al.*, Nucleic Acids Res. 15: 6131-6148 (1987)) or a chimeric RNA-DNA analogue (Inoue *et al.*, FEBS Lett. 215: 327-330 (1987)). Antisense nucleic acids sometimes are composed of DNA or PNA or any other nucleic acid derivatives described previously.

[0139] In another embodiment, an antisense nucleic acid is a ribozyme. A ribozyme having specificity for a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence can include one or more sequences complementary to such a nucleotide sequence, and a sequence having a known catalytic region responsible for mRNA cleavage (see *e.g.*, U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, Nature 334: 585-591 (1988)). For example, a derivative of a Tetrahymena L-19 IVS RNA is sometimes utilized in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA (see *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742). Also, target mRNA sequences can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see *e.g.*, Bartel & Szostak, Science 261: 1411-1418 (1993)).

[0140] Breast cancer directed molecules include in certain embodiments nucleic acids that can form triple helix structures with a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence or a substantially identical sequence thereof, especially one that includes a regulatory region that controls expression of a polypeptide. Gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence or a substantially identical sequence (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of a gene in target cells (see *e.g.*, Helene, Anticancer Drug Des. 6(6): 569-84 (1991); Helene *et al.*, Ann. N.Y. Acad. Sci. 660: 27-36 (1992); and Maher, Bioassays 14(12): 807-15 (1992). Potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0141] Breast cancer directed molecules include RNAi and siRNA nucleic acids. Gene expression may be inhibited by the introduction of double-stranded RNA (dsRNA), which induces

potent and specific gene silencing, a phenomenon called RNA interference or RNAi. See, *e.g.*, Fire *et al.*, US Patent Number 6,506,559; Tuschl *et al.* PCT International Publication No. WO 01/75164; Kay *et al.* PCT International Publication No. WO 03/010180A1; or Bosher JM, Labouesse, Nat Cell Biol 2000 Feb;2(2):E31-6. This process has been improved by decreasing the size of the double-stranded RNA to 20-24 base pairs (to create small-interfering RNAs or siRNAs) that "switched off" genes in mammalian cells without initiating an acute phase response, i.e., a host defense mechanism that often results in cell death (see, *e.g.*, Caplen *et al.* Proc Natl Acad Sci U S A. 2001 Aug 14;98(17):9742-7 and Elbashir *et al.* Methods 2002 Feb;26(2):199-213). There is increasing evidence of post-transcriptional gene silencing by RNA interference (RNAi) for inhibiting targeted expression in mammalian cells at the mRNA level, in human cells. There is additional evidence of effective methods for inhibiting the proliferation and migration of tumor cells in human patients, and for inhibiting metastatic cancer development (see, *e.g.*, U.S. Patent Application No. US2001000993183; Caplen *et al.* Proc Natl Acad Sci U S A; and Abderrahmani *et al.* Mol Cell Biol 2001 Nov21(21):7256-67).

[0142] An "siRNA" or "RNAi" refers to a nucleic acid that forms a double stranded RNA and has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is delivered to or expressed in the same cell as the gene or target gene. "siRNA" refers to short double-stranded RNA formed by the complementary strands. Complementary portions of the siRNA that hybridize to form the double stranded molecule often have substantial or complete identity to the target molecule sequence. In one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA.

[0143] When designing the siRNA molecules, the targeted region often is selected from a given DNA sequence beginning 50 to 100 nucleotides downstream of the start codon. See, *e.g.*, Elbashir et al,. Methods 26:199-213 (2002). Initially, 5' or 3' UTRs and regions nearby the start codon were avoided assuming that UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. Sometimes regions of the target 23 nucleotides in length conforming to the sequence motif AA(N19)TT (N, an nucleotide), and regions with approximately 30% to 70% G/C-content (often about 50% G/C-content) often are selected. If no suitable sequences are found, the search often is extended using the motif NA(N21). The sequence of the sense siRNA sometimes corresponds to (N19) TT or N21 (position 3 to 23 of the 23-nt motif), respectively. In the latter case, the 3' end of the sense siRNA often is converted to TT. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. The antisense siRNA is synthesized as the complement to position 1 to 21 of the 23-nt motif. Because position 1 of the 23-nt motif is not recognized sequence-specifically by the antisense siRNA, the 3'-most nucleotide residue of the antisense siRNA can be chosen deliberately. However, the penultimate nucleotide of the antisense siRNA (complementary to position 2 of the 23-nt motif) often is complementary to the targeted

sequence. For simplifying chemical synthesis, TT often is utilized. siRNAs corresponding to the target motif NAR(N17)YNN, where R is purine (A,G) and Y is pyrimidine (C,U), often are selected. Respective 21 nucleotide sense and antisense siRNAs often begin with a purine nucleotide and can also be expressed from pol III expression vectors without a change in targeting site. Expression of RNAs from pol III promoters often is efficient when the first transcribed nucleotide is a purine.

[0144] The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Often, the siRNA is about 15 to about 50 nucleotides in length (*e.g.*, each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, sometimes about 20-30 nucleotides in length or about 20-25 nucleotides in length, *e.g.*, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. The siRNA sometimes is about 21 nucleotides in length. Methods of using siRNA are well known in the art, and specific siRNA molecules may be purchased from a number of companies including Dharmacon Research, Inc.

[0145] Antisense, ribozyme, RNAi and siRNA nucleic acids can be altered to form modified nucleic acid molecules. The nucleic acids can be altered at base moieties, sugar moieties or phosphate backbone moieties to improve stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup *et al.*, Bioorganic & Medicinal Chemistry 4 (1): 5-23 (1996)). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic such as a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. Synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described, for example, in Hyrup *et al.*, (1996) supra and Perry-O'Keefe *et al.*, Proc. Natl. Acad. Sci. 93: 14670-675 (1996).

[0146] PNA nucleic acids can be used in prognostic, diagnostic, and therapeutic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNA nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as "artificial restriction enzymes" when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup (1996) supra)); or as probes or primers for DNA sequencing or hybridization (Hyrup *et al.*, (1996) supra; Perry-O'Keefe supra).

[0147] In other embodiments, oligonucleotides may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across cell membranes (see *e.g.*, Letsinger *et al.*, Proc. Natl. Acad. Sci. USA 86: 6553-6556 (1989); Lemaitre *et al.*, Proc. Natl. Acad. Sci. USA 84: 648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.*, Bio-Techniques 6: 958-

976 (1988)) or intercalating agents. (See, *e.g.*, Zon, Pharm. Res. 5: 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0148] Also included herein are molecular beacon oligonucleotide primer and probe molecules having one or more regions complementary to a nucleotide sequence of SEQ ID NO: 1-8 or a substantially identical sequence thereof, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantifying the presence of the nucleic acid in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.*, U.S. Patent No. 5,854,033; Nazarenko *et al.*, U.S. Patent No. 5,866,336, and Livak *et al.*, U.S. Patent 5,876,930.

Antibodies

[0149] The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. An antibody sometimes is a polyclonal, monoclonal, recombinant (*e.g.*, a chimeric or humanized), fully human, non-human (*e.g.*, murine), or a single chain antibody. An antibody may have effector function and can fix complement, and is sometimes coupled to a toxin or imaging agent.

[0150] A full-length polypeptide or antigenic peptide fragment encoded by a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleotide sequence can be used as an immunogen or can be used to identify antibodies made with other immunogens, *e.g.*, cells, membrane preparations, and the like. An antigenic peptide often includes at least 8 amino acid residues of the amino acid sequences encoded by a nucleotide sequence of SEQ ID NO: 1-8, or substantially identical sequence thereof, and encompasses an epitope. Antigenic peptides sometimes include 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, or 30 or more amino acids. Hydrophilic and hydrophobic fragments of polypeptides sometimes are used as immunogens.

[0151] Epitopes encompassed by the antigenic peptide are regions located on the surface of the polypeptide (*e.g.*, hydrophilic regions) as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human polypeptide sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the polypeptide and are thus likely to constitute surface residues useful for targeting antibody production. The antibody may bind an epitope on any domain or region on polypeptides described herein.

[0152] Also, chimeric, humanized, and completely human antibodies are useful for applications which include repeated administration to subjects. Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant

DNA techniques known in the art, for example using methods described in Robinson et al International Application No. PCT/US86/02269; Akira, et al European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al European Patent Application 173,494; Neuberger et al PCT International Publication No. WO 86/01533; Cabilly et al U.S. Patent No. 4,816,567; Cabilly et al European Patent Application 125,023; Better *et al.*, Science 240: 1041-1043 (1988); Liu *et al.*, Proc. Natl. Acad. Sci. USA 84: 3439-3443 (1987); Liu *et al.*, J. Immunol. 139: 3521-3526 (1987); Sun *et al.*, Proc. Natl. Acad. Sci. USA 84: 214-218 (1987); Nishimura *et al.*, Canc. Res. 47: 999-1005 (1987); Wood *et al.*, Nature 314: 446-449 (1985); and Shaw *et al.*, J. Natl. Cancer Inst. 80: 1553-1559 (1988); Morrison, S. L., Science 229: 1202-1207 (1985); Oi *et al.*, BioTechniques 4: 214 (1986); Winter U.S. Patent 5,225,539; Jones *et al.*, Nature 321: 552-525 (1986); Verhoeyan *et al.*, Science 239: 1534; and Beidler *et al.*, J. Immunol. 141: 4053-4060 (1988).

[0153] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar, Int. Rev. Immunol. 13: 65-93 (1995); and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above. Completely human antibodies that recognize a selected epitope also can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody (*e.g.*, a murine antibody) is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described for example by Jespers *et al.*, Bio/Technology 12: 899-903 (1994).

[0154] Antibody can be a single chain antibody. A single chain antibody (scFV) can be engineered (see, *e.g.*, Colcher *et al.*, Ann. N Y Acad. Sci. 880: 263-80 (1999); and Reiter, Clin. Cancer Res. 2: 245-52 (1996)). Single chain antibodies can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target polypeptide.

[0155] Antibodies also may be selected or modified so that they exhibit reduced or no ability to bind an Fc receptor. For example, an antibody may be an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor (*e.g.*, it has a mutagenized or deleted Fc receptor binding region).

[0156] Also, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1 dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and

analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thiotepa chlorambucil, melphalan, carmustine (BCNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

[0157] Antibody conjugates can be used for modifying a given biological response. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, ?-interferon, a-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Also, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980, for example.

[0158] An antibody (*e.g.*, monoclonal antibody) can be used to isolate target polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an antibody can be used to detect a target polypeptide (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, *e.g.*, to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, ß-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H. Also, an antibody can be utilized as a test molecule for determining whether it can treat breast cancer, and as a therapeutic for administration to a subject for treating breast cancer.

[0159] An antibody can be made by immunizing with a purified antigen, or a fragment thereof, *e.g.*, a fragment described herein, a membrane associated antigen, tissues, *e.g.*, crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[0160] Included herein are antibodies which bind only a native polypeptide, only denatured or otherwise non-native polypeptide, or which bind both, as well as those having linear or conformational epitopes. Conformational epitopes sometimes can be identified by selecting antibodies that bind to native but not denatured polypeptide. Also featured are antibodies that specifically bind to a polypeptide variant associated with breast cancer.

Screening Assays

[0161] Featured herein are methods for identifying a candidate therapeutic for treating breast cancer. The methods comprise contacting a test molecule with a target molecule in a system. A "target molecule" as used herein refers to a nucleic acid of SEQ ID NO: 1-8, a substantially identical nucleic acid thereof, or a fragment thereof, and an encoded polypeptide of the foregoing. The method also comprises determining the presence or absence of an interaction between the test molecule and the target molecule, where the presence of an interaction between the test molecule and the nucleic acid or polypeptide identifies the test molecule as a candidate breast cancer therapeutic. The interaction between the test molecule and the target molecule may be quantified.

[0162] Test molecules and candidate therapeutics include, but are not limited to, compounds, antisense nucleic acids, siRNA molecules, ribozymes, polypeptides or proteins encoded by a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acids, or a substantially identical sequence or fragment thereof, and immunotherapeutics (*e.g.*, antibodies and HLA-presented polypeptide fragments). A test molecule or candidate therapeutic may act as a modulator of target molecule concentration or target molecule function in a system. A "modulator" may agonize (i.e., up-regulates) or antagonize (i.e., down-regulates) a target molecule concentration partially or completely in a system by affecting such cellular functions as DNA replication and/or DNA processing (*e.g.*, DNA methylation or DNA repair), RNA transcription and/or RNA processing (*e.g.*, removal of intronic sequences and/or translocation of spliced mRNA from the nucleus), polypeptide production (*e.g.*, translation of the polypeptide from mRNA), and/or polypeptide post-translational modification (*e.g.*, glycosylation, phosphorylation, and proteolysis of pro-polypeptides). A modulator may also agonize or antagonize a biological function of a target molecule partially or completely, where the function may include adopting a certain structural conformation, interacting with one or more binding partners, ligand binding, catalysis (*e.g.*, phosphorylation, dephosphorylation, hydrolysis, methylation, and isomerization), and an effect upon a cellular event (*e.g.*, effecting progression of breast cancer).

[0163] As used herein, the term "system" refers to a cell free *in vitro* environment and a cell-based environment such as a collection of cells, a tissue, an organ, or an organism. A system is "contacted" with a test molecule in a variety of manners, including adding molecules in solution and allowing them to interact with one another by diffusion, cell injection, and any administration routes in an animal. As used herein, the term "interaction" refers to an effect of a test molecule on test

molecule, where the effect sometimes is binding between the test molecule and the target molecule, and sometimes is an observable change in cells, tissue, or organism.

[0164] There are many standard methods for detecting the presence or absence of an interaction between a test molecule and a target molecule. For example, titrametric, acidimetric, radiometric, NMR, monolayer, polarographic, spectrophotometric, fluorescent, and ESR assays probative of a target molecule interaction may be utilized.

[0165] In general, an interaction can be determined by labeling the test molecule and/or the *DLG1, KIAA0783, DPF3* or *CENPC1* molecule, where the label is covalently or non-covalently attached to the test molecule or *DLG1, KIAA0783, DPF3* or *CENPC1* molecule. The label is sometimes a radioactive molecule such as ^{125}I, ^{131}I, ^{35}S or ^{3}H, which can be detected by direct counting of radioemission or by scintillation counting. Also, enzymatic labels such as horseradish peroxidase, alkaline phosphatase, or luciferase may be utilized where the enzymatic label can be detected by determining conversion of an appropriate substrate to product. Also, presence or absence of an interaction can be determined without labeling. For example, a microphysiometer (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indication of an interaction between a test molecule and *DLG1, KIAA0783, DPF3* or *CENPC1* (McConnell, H. M. et al., Science 257: 1906-1912 (1992)).

[0166] In cell-based systems, cells typically include a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid or polypeptide or variants thereof and are often of mammalian origin, although the cell can be of any origin. Whole cells, cell homogenates, and cell fractions (e.g., cell membrane fractions) can be subjected to analysis. Where interactions between a test molecule with a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or variant thereof are monitored, soluble and/or membrane bound forms of the polypeptide or variant may be utilized. Where membrane-bound forms of the polypeptide are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0167] An interaction between two molecules also can be detected by monitoring fluorescence energy transfer (FET) (see, for example, Lakowicz et al., U.S. Patent No. 5,631,169; Stavrianopoulos et al. U.S. Patent No. 4,868,103). A fluorophore label on a first, "donor" molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, "acceptor" molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the "donor" polypeptide molecule may simply utilize the natural fluorescent energy of tryptophan residues.

Labels are chosen that emit different wavelengths of light, such that the "acceptor" molecule label may be differentiated from that of the "donor". Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the "acceptor" molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

[0168] In another embodiment, determining the presence or absence of an interaction between a test molecule and a *DLG1, KIAA0783, DPF3* or *CENPC1* molecule can be effected by using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander & Urbaniczk, Anal. Chem. 63: 2338-2345 (1991) and Szabo et al., Curr. Opin. Struct. Biol. 5: 699-705 (1995)). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0169] In another embodiment, the *DLG1, KIAA0783, DPF3* or *CENPC1* molecule or test molecules are anchored to a solid phase. The *DLG1, KIAA0783, DPF3* or *CENPC1* molecule/test molecule complexes anchored to the solid phase can be detected at the end of the reaction. The target *DLG1, KIAA0783, DPF3* or *CENPC1* molecule is often anchored to a solid surface, and the test molecule, which is not anchored, can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0170] It may be desirable to immobilize a *DLG1, KIAA0783, DPF3* or *CENPC1* molecule, an anti-*DLG1, KIAA0783, DPF3* or *CENPC1* antibody, or test molecules to facilitate separation of complexed from uncomplexed forms of *DLG1, KIAA0783, DPF3* or *CENPC1* molecules and test molecules, as well as to accommodate automation of the assay. Binding of a test molecule to a *DLG1, KIAA0783, DPF3* or *CENPC1* molecule can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion polypeptide can be provided which adds a domain that allows a *DLG1, KIAA0783, DPF3* or *CENPC1* molecule to be bound to a matrix. For example, glutathione-S-transferase/*DLG1, KIAA0783, DPF3* or *CENPC1* fusion polypeptides or glutathione-S-transferase/target fusion polypeptides can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivitized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target polypeptide or *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix

immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of *DLG1, KIAA0783, DPF3* or *CENPC1* binding or activity determined using standard techniques.

[0171] Other techniques for immobilizing a *DLG1, KIAA0783, DPF3* or *CENPC1* molecule on matrices include using biotin and streptavidin. For example, biotinylated *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

[0172] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

[0173] In one embodiment, this assay is performed utilizing antibodies reactive with *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or test molecules but which do not interfere with binding of the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide to its test molecule. Such antibodies can be derivitized to the wells of the plate, and unbound target or *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or test molecule.

[0174] Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A. P., Trends Biochem Sci Aug;18(8): 284-7 (1993)); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel et al., eds. Current Protocols in Molecular Biology , J. Wiley: New York (1999)); and immunoprecipitation (see, for example, Ausubel, F. et al., eds. Current Protocols in Molecular Biology , J. Wiley: New York (1999)). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, J Mol. Recognit. Winter; 11(1-6): 141-8 (1998); Hage & Tweed, J. Chromatogr. B Biomed. Sci. Appl. Oct 10; 699 (1-2): 499-525 (1997)). Further, fluorescence energy transfer may

also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0175] In another embodiment, modulators of *DLG1, KIAA0783, DPF3* or *CENPC1* expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide evaluated relative to the level of expression of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide in the absence of the candidate compound. When expression of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide expression. Alternatively, when expression of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide expression. The level of *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide expression can be determined by methods described herein for detecting *DLG1, KIAA0783, DPF3* or *CENPC1* mRNA or polypeptide.

[0176] In another embodiment, binding partners that interact with a *DLG1, KIAA0783, DPF3* or *CENPC1* molecule are detected. The *DLG1, KIAA0783, DPF3* or *CENPC1* molecules can interact with one or more cellular or extracellular macromolecules, such as polypeptides, in vivo, and these molecules that interact with *DLG1, KIAA0783, DPF3* or *CENPC1* molecules are referred to herein as "binding partners." Molecules that disrupt such interactions can be useful in regulating the activity of the target gene product. Such molecules can include, but are not limited to molecules such as antibodies, peptides, and small molecules. Target genes/products for use in this embodiment often are the *DLG1, KIAA0783, DPF3* or *CENPC1* genes herein identified. In an alternative embodiment, provided is a method for determining the ability of the test compound to modulate the activity of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide through modulation of the activity of a downstream effector of a *DLG1, KIAA0783, DPF3* or *CENPC1* target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[0177] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), e.g., a substrate, a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex

in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases where it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[0178] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

[0179] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

[0180] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[0181] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect

anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[0182] In an alternate embodiment, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[0183] Also, binding partners of *DLG1, KIAA0783, DPF3* or *CENPC1* molecules can be identified in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al., Cell 72:223-232 (1993); Madura et al., J. Biol. Chem. 268: 12046-12054 (1993); Bartel et al., Biotechniques 14: 920-924 (1993); Iwabuchi et al., Oncogene 8: 1693-1696 (1993); and Brent WO94/10300), to identify other polypeptides, which bind to or interact with *DLG1, KIAA0783, DPF3* or *CENPC1* ("*DLG1, KIAA0783, DPF3* or *CENPC1*-binding polypeptides" or "*DLG1, KIAA0783, DPF3* or *CENPC1*-bp") and are involved in *DLG1, KIAA0783, DPF3* or *CENPC1* activity. Such *DLG1, KIAA0783, DPF3* or *CENPC1*-bps can be activators or inhibitors of signals by the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptides or *DLG1, KIAA0783, DPF3* or *CENPC1* targets as, for example, downstream elements of a *DLG1, KIAA0783, DPF3* or *CENPC1*-mediated signaling pathway.

[0184] A two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified polypeptide ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the: *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide can be the fused to the activator domain.) If the "bait" and the "prey" polypeptides are able to interact, in vivo, forming a *DLG1, KIAA0783, DPF3* or *CENPC1*-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the polypeptide which interacts with the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide.

[0185] Candidate therapeutics for treating breast cancer are identified from a group of test molecules that interact with a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid or polypeptide. Test molecules are normally ranked according to the degree with which they interact or modulate (e.g., agonize or antagonize) DNA replication and/or processing, RNA transcription and/or processing, polypeptide production and/or processing, and/or function of *DLG1, KIAA0783, DPF3* or *CENPC1* molecules, for example, and then top ranking modulators are selected. In a preferred embodiment, the candidate therapeutic (i.e., test molecule) acts as a *DLG1, KIAA0783, DPF3* or *CENPC1* antagonist. Also, pharmacogenomic information described herein can determine the rank of a modulator. Candidate therapeutics typically are formulated for administration to a subject.

Therapeutic Treatments

[0186] Formulations or pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier, a compound, an antisense nucleic acid, a ribozyme, an antibody, a binding partner that interacts with a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide, a *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid, or a fragment thereof. The formulated molecule may be one that is identified by a screening method described above. Also, formulations may comprise a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or fragment thereof. As used herein, the term "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[0187] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0188] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be

prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0189] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride sometimes are included in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0190] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation often utilized are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0191] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

[0192] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are

used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Molecules can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0193] In one embodiment, active molecules are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0194] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0195] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Molecules which exhibit high therapeutic indices often are utilized. While molecules that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0196] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such molecules often lies within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any molecules used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the

test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0197] As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, sometimes about 0.01 to 25 mg/kg body weight, often about 0.1 to 20 mg/kg body weight, and more often about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, sometimes between 2 to 8 weeks, often between about 3 to 7 weeks, and more often for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment, or sometimes can include a series of treatments.

[0198] With regard to polypeptide formulations, featured herein is a method for treating breast cancer in a subject, which comprises contacting one or more cells in the subject with a first *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide, where the subject comprises a second *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide having one or more polymorphic variations associated with cancer, and where the first polypeptide comprises fewer polymorphic variations associated with cancer than the second polypeptide. The first and second polypeptides are encoded by a nucleic acid which comprises a nucleotide sequence selected from the group consisting of the nucleotide sequence of SEQ ID NO: 1-8; a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-8; a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-8 and a nucleotide sequence 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-8. The subject is often a human.

[0199] For antibodies, a dosage of 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg) is often utilized. If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is often appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al., J. Acquired Immune Deficiency Syndromes and Human Retrovirology 14:193 (1997).

[0200] Antibody conjugates can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins

may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0201] For compounds, exemplary doses include milligram or microgram amounts of the compound per kilogram of subject or sample weight, for example, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid described herein, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0202] *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid molecules can be inserted into vectors and used in gene therapy methods for treating breast cancer. Featured herein is a method for treating breast cancer in a subject, which comprises contacting one or more cells in the subject with a first *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid, where genomic DNA in the subject comprises a second *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid comprising one or more polymorphic variations associated with breast cancer, and where the first *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid comprises fewer polymorphic variations associated with breast cancer. The first and second nucleic acids typically comprise a nucleotide sequence selected from the group consisting of the nucleotide sequence of SEQ ID NO: 1-8; a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-8; a nucleotide sequence that is 90% or more identical to the nucleotide sequence of SEQ ID NO: 1-8, and a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-8. The subject often is a human.

[0203] Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen et al., (1994) Proc. Natl. Acad. Sci. USA 91:3054-3057). Pharmaceutical preparations of gene therapy

vectors can include a gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells (e.g., retroviral vectors) the pharmaceutical preparation can include one or more cells which produce the gene delivery system. Examples of gene delivery vectors are described herein.

[0204] Pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0205] Pharmaceutical compositions of active ingredients can be administered by any of the paths described herein for therapeutic and prophylactic methods for treating breast cancer. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from pharmacogenomic analyses described herein. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

[0206] Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the *DLG1, KIAA0783, DPF3* or *CENPC1* aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of *DLG1, KIAA0783, DPF3* or *CENPC1* aberrance, for example, a *DLG1, KIAA0783, DPF3* or *CENPC1* molecule, *DLG1, KIAA0783, DPF3* or *CENPC1* agonist, or *DLG1, KIAA0783, DPF3* or *CENPC1* antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

[0207] As discussed, successful treatment of *DLG1, KIAA0783, DPF3* or *CENPC1* disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds (e.g., an agent identified using an assays described above) that exhibit negative modulatory activity can be used to prevent and/or treat breast cancer. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and FAb, F(ab')2 and FAb expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[0208] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[0209] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances where the target gene encodes an extracellular polypeptide, normal target gene polypeptide often is co-administered into the cell or tissue to maintain the requisite level of cellular or tissue target gene activity.

[0210] Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by *DLG1, KIAA0783, DPF3* or *CENPC1* expression is through the use of aptamer molecules specific for *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to polypeptide ligands (see, e.g., Osborne, et al., Curr. Opin. Chem. Biol.1(1): 5-9 (1997); and Patel, D. J., Curr. Opin. Chem. Biol. Jun;1(1): 32-46 (1997)). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic polypeptide molecules may be, aptamers offer a method by which *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

[0211] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of *DLG1, KIAA0783, DPF3* or *CENPC1* disorders. For a description of antibodies, see the Antibody section above.

[0212] In circumstances where injection of an animal or a human subject with a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against *DLG1, KIAA0783, DPF3* or *CENPC1* through the use of anti-idiotypic antibodies (see, for example, Herlyn, D., Ann. Med.;31(1): 66-78 (1999); and Bhattacharya-Chatterjee & Foon, Cancer Treat. Res.; 94: 51-68 (1998)). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide. Vaccines directed to a disease characterized by *DLG1, KIAA0783, DPF3* or *CENPC1* expression may also be generated in this fashion.

[0213] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be utilized. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen often is utilized. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody

can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco et al., Proc. Natl. Acad. Sci. USA 90: 7889-7893 (1993)).

[0214] *DLG1, KIAA0783, DPF3* or *CENPC1* molecules and compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate *DLG1, KIAA0783, DPF3* or *CENPC1* disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

[0215] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit large therapeutic indices often are utilized. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0216] Data obtained from cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds often lies within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in a method described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0217] Another example of effective dose determination for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate *DLG1, KIAA0783, DPF3* or *CENPC1* activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell et al., Current Opinion in Biotechnology 7: 89-94 (1996) and in Shea, Trends in Polymer Science 2: 166-173 (1994). Such "imprinted" affinity matrixes are amenable to ligand-

binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, et al., Nature 361: 645-647 (1993). Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of *DLG1, KIAA0783, DPF3* or *CENPC1* can be readily monitored and used in calculations of IC_{50}. Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50}. A rudimentary example of such a "biosensor" is discussed in Kriz et al., Analytical Chemistry 67: 2142-2144 (1995).

[0218] Provided herein are methods of modulating *DLG1, KIAA0783, DPF3* or *CENPC1* expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method involves contacting a cell with a *DLG1, KIAA0783, DPF3* or *CENPC1* or agent that modulates one or more of the activities of *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide activity associated with the cell. An agent that modulates *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide activity can be an agent as described herein, such as a nucleic acid or a polypeptide, a naturally-occurring target molecule of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide (e.g., a *DLG1, KIAA0783, DPF3* or *CENPC1* substrate or receptor), a *DLG1, KIAA0783, DPF3* or *CENPC1* antibody, a *DLG1, KIAA0783, DPF3* or *CENPC1* agonist or antagonist, a peptidomimetic of a *DLG1, KIAA0783, DPF3* or *CENPC1* agonist or antagonist, or other small molecule.

[0219] In one embodiment, the agent stimulates one or more *DLG1, KIAA0783, DPF3* or *CENPC1* activities. Examples of such stimulatory agents include active *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide and a nucleic acid molecule encoding *DLG1, KIAA0783, DPF3* or *CENPC1*. In another embodiment, the agent inhibits one or more *DLG1, KIAA0783, DPF3* or *CENPC1* activities. Examples of such inhibitory agents include antisense *DLG1, KIAA0783, DPF3* or *CENPC1* nucleic acid molecules, anti-*DLG1, KIAA0783, DPF3* or *CENPC1* antibodies, and *DLG1, KIAA0783, DPF3* or *CENPC1* inhibitors. These modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g., by administering the agent to a subject). As such, provided are methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a *DLG1, KIAA0783, DPF3* or *CENPC1* polypeptide or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) *DLG1, KIAA0783, DPF3* or *CENPC1* expression or activity. In a preferred embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that inhibits *DLG1, KIAA0783, DPF3* or *CENPC1* expression or activity. In another embodiment, the method involves administering a *DLG1,*

KIAA0783, DPF3 or *CENPC1* polypeptide or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted *DLG1, KIAA0783, DPF3* or *CENPC1* expression or activity.

[0220] Stimulation of *DLG1, KIAA0783, DPF3* or *CENPC1* activity is desirable in situations in which *DLG1, KIAA0783, DPF3* or *CENPC1* is abnormally downregulated and/or in which increased *DLG1, KIAA0783, DPF3* or *CENPC1* activity is likely to have a beneficial effect. For example, stimulation of *DLG1, KIAA0783, DPF3* or *CENPC1* activity is desirable in situations in which a *DLG1, KIAA0783, DPF3* or *CENPC1* is downregulated and/or in which increased *DLG1, KIAA0783, DPF3* or *CENPC1* activity is likely to have a beneficial effect. Likewise, inhibition of *DLG1, KIAA0783, DPF3* or *CENPC1* activity is desirable in situations in which *DLG1, KIAA0783, DPF3* or *CENPC1* is abnormally upregulated and/or in which decreased *DLG1, KIAA0783, DPF3* or *CENPC1* activity is likely to have a beneficial effect.

Methods of Treatment

[0221] In another aspect, provided are methods for identifying a risk of cancer in an individual as described herein and, if a genetic predisposition is identified, treating that individual to delay or reduce or prevent the development of cancer. Such a procedure can be used to treat breast cancer. Optionally, treating an individual for cancer may include inhibiting cellular proliferation, inhibiting metastasis, inhibiting invasion, or preventing tumor formation or growth as defined herein. Suitable treatments to prevent or reduce or delay breast cancer focus on inhibiting additional cellular proliferation, inhibiting metastasis, inhibiting invasion, and preventing further tumor formation or growth. Treatment usually includes surgery followed by radiation therapy. Surgery may be a lumpectomy or a mastectomy (e.g., total, simple or radical). Even if the doctor removes all of the cancer that can be seen at the time of surgery, the patient may be given radiation therapy, chemotherapy, or hormone therapy after surgery to try to kill any cancer cells that may be left. Radiation therapy is the use of x-rays or other types of radiation to kill cancer cells and shrink tumors. Radiation therapy may use external radiation (using a machine outside the body) or internal radiation. Chemotherapy is the use of drugs to kill cancer cells. Chemotherapy may be taken by mouth, or it may be put into the body by inserting a needle into a vein or muscle. Hormone therapy often focuses on estrogen and progesterone, which are hormones that affect the way some cancers grow. If tests show that the cancer cells have estrogen and progesterone receptors (molecules found in some cancer cells to which estrogen and progesterone will attach), hormone therapy is used to block the way these hormones help the cancer grow. Hormone therapy with tamoxifen is often given to patients with early stages of breast cancer and those with metastatic breast cancer. Other types of treatment being tested in clinical trials include sentinel lymph node biopsy followed by surgery and high-dose chemotherapy with bone marrow transplantation and peripheral blood stem cell transplantation. Any preventative/therapeutic treatment known in the art may be prescribed and/or administered, including, for example, surgery, chemotherapy and/or radiation treatment, and any of the treatments may be used

in combination with one another to treat or prevent breast cancer (e.g., surgery followed by radiation therapy).

[0222] Also provided are methods of preventing or treating cancer comprising providing an individual in need of such treatment with a *DLG1, KIAA0783, DPF3* or *CENPC1* inhibitor that reduces or inhibits the overexpression of mutant *DLG1, KIAA0783, DPF3* or *CENPC1* (e.g., a *DLG1, KIAA0783, DPF3* or *CENPC1* polynucleotide with an allele that is associated with cancer). Included herein are methods of reducing or blocking the expression of *DLG1, KIAA0783, DPF3* or *CENPC1* comprising providing or administering to individuals in need of reducing or blocking the expression of *DLG1, KIAA0783, DPF3* or *CENPC1* a pharmaceutical or physiologically acceptable composition comprising a molecule capable of inhibiting expression of *DLG1, KIAA0783, DPF3* or *CENPC1*, e.g., a siRNA molecule. Also included herein are methods of reducing or blocking the expression of secondary regulatory genes regulated by *DLG1, KIAA0783, DPF3* or *CENPC1* that play a role in oncogenesis which comprises introducing competitive inhibitors that target *DLG1, KIAA0783, DPF3* or *CENPC1*'s effect on these regulatory genes or that block the binding of positive factors necessary for the expression of these regulatory genes.

[0223] The examples set forth below are intended to illustrate but not limit the invention.

Examples

[0224] In the following studies a group of subjects were selected according to specific parameters relating to breast cancer. Nucleic acid samples obtained from individuals in the study group were subjected to genetic analysis, which identified associations between breast cancer and certain polymorphic regions in the *DLG1, KIAA0783, DPF3* and *CENPC1* genes (herein referred to as "target genes", "target nucleotides", "target polypeptides" or simply "targets"). Methods are described for producing DLG1, KIAA0783, DPF3 and CENPC1 polypeptides and polypeptide variants *in vitro* or *in vivo*. DLG1, KIAA0783, DPF3 and CENPC1 nucleic acids or polypeptides and variants thereof are utilized for screening test molecules for those that interact with DLG1, KIAA0783, DPF3 and CENPC1 molecules. Test molecules identified as interactors with DLG1, KIAA0783, DPF3 and CENPC1 molecules and variants are further screened *in vivo* to determine whether they treat breast cancer.

Example 1

Samples and Pooling Strategies

Sample Selection

[0225] Blood samples were collected from individuals diagnosed with breast cancer, which were referred to as case samples. Also, blood samples were collected from individuals not diagnosed with breast cancer as gender and age-matched controls. All of the samples were of German/German

descent. A database was created that listed all phenotypic trait information gathered from individuals for each case and control sample. Genomic DNA was extracted from each of the blood samples for genetic analyses.

DNA Extraction from Blood Samples

[0226] Six to ten milliliters of whole blood was transferred to a 50 ml tube containing 27 ml of red cell lysis solution (RCL). The tube was inverted until the contents were mixed. Each tube was incubated for 10 minutes at room temperature and inverted once during the incubation. The tubes were then centrifuged for 20 minutes at 3000 x g and the supernatant was carefully poured off. 100-200 µl of residual liquid was left in the tube and was pipetted repeatedly to resuspend the pellet in the residual supernatant. White cell lysis solution (WCL) was added to the tube and pipetted repeatedly until completely mixed. While no incubation was normally required, the solution was incubated at 37°C or room temperature if cell clumps were visible after mixing until the solution was homogeneous. 2 ml of protein precipitation was added to the cell lysate. The mixtures were vortexed vigorously at high speed for 20 sec to mix the protein precipitation solution uniformly with the cell lysate, and then centrifuged for 10 minutes at 3000 x g. The supernatant containing the DNA was then poured into a clean 15 ml tube, which contained 7 ml of 100% isopropanol. The samples were mixed by inverting the tubes gently until white threads of DNA were visible. Samples were centrifuged for 3 minutes at 2000 x g and the DNA was visible as a small white pellet. The supernatant was decanted and 5 ml of 70% ethanol was added to each tube. Each tube was inverted several times to wash the DNA pellet, and then centrifuged for 1 minute at 2000 x g. The ethanol was decanted and each tube was drained on clean absorbent paper. The DNA was dried in the tube by inversion for 10 minutes, and then 1000 µl of 1X TE was added. The size of each sample was estimated, and less TE buffer was added during the following DNA hydration step if the sample was smaller. The DNA was allowed to rehydrate overnight at room temperature, and DNA samples were stored at 2-8°C.

[0227] DNA was quantified by placing samples on a hematology mixer for at least 1 hour. DNA was serially diluted (typically 1:80, 1:160, 1:320, and 1:640 dilutions) so that it would be within the measurable range of standards. 125 µl of diluted DNA was transferred to a clear U-bottom microtitre plate, and 125 µl of 1X TE buffer was transferred into each well using a multichannel pipette. The DNA and 1X TE were mixed by repeated pipetting at least 15 times, and then the plates were sealed. 50 µl of diluted DNA was added to wells A5-H12 of a black flat bottom microtitre plate. Standards were inverted six times to mix them, and then 50 µl of 1X TE buffer was pipetted into well A1, 1000 ng/ml of standard was pipetted into well A2, 500 ng/ml of standard was pipetted into well A3, and 250 ng/ml of standard was pipetted into well A4. PicoGreen (Molecular Probes, Eugene, Oregon) was thawed and freshly diluted 1:200 according to the number of plates that were being measured.

PicoGreen was vortexed and then 50µl was pipetted into all wells of the black plate with the diluted DNA. DNA and PicoGreen were mixed by pipetting repeatedly at least 10 times with the multichannel pipette. The plate was placed into a Fluoroskan Ascent Machine (microplate fluorometer produced by Labsystems) and the samples were allowed to incubate for 3 minutes before the machine was run using filter pairs 485 nm excitation and 538 nm emission wavelengths. Samples having measured DNA concentrations of greater than 450 ng/µl were re-measured for conformation. Samples having measured DNA concentrations of 20 ng/µl or less were re-measured for confirmation.

Pooling Strategies

[0228] Samples were placed into one of two groups based on disease status. The two groups were female case groups and female control groups. A select set of samples from each group were utilized to generate pools, and one pool was created for each group. Each individual sample in a pool was represented by an equal amount of genomic DNA. For example, where 25 ng of genomic DNA was utilized in each PCR reaction and there were 200 individuals in each pool, each individual would provide 125 pg of genomic DNA. Inclusion or exclusion of samples for a pool was based upon the following criteria: the sample was derived from an individual characterized as Caucasian; the sample was derived from an individual of German paternal and maternal descent; the database included relevant phenotype information for the individual; case samples were derived from individuals diagnosed with breast cancer; control samples were derived from individuals free of cancer and no family history of breast cancer; and sufficient genomic DNA was extracted from each blood sample for all allelotyping and genotyping reactions performed during the study. Phenotype information included pre- or post-menopausal, familial predisposition, country or origin of mother and father, diagnosis with breast cancer (date of primary diagnosis, age of individual as of primary diagnosis, grade or stage of development, occurrence of metastases, e.g., lymph node metastases, organ metastases), condition of body tissue (skin tissue, breast tissue, ovary tissue, peritoneum tissue and myometrium), method of treatment (surgery, chemotherapy, hormone therapy, radiation therapy). Samples that met these criteria were added to appropriate pools based on gender and disease status.

[0229] The selection process yielded the pools set forth in Table 1, which were used in the studies that follow:

Table 1

	Female CASE	**Female CONTROL**
Pool size (Number)	272	276
Pool Criteria (ex: case/control)	case	control

Mean Age (ex: years)	59.6	55.4

Example 2

Association of Polymorphic Variants with Breast cancer

[0230] A whole-genome screen was performed to identify particular SNPs associated with occurrence of breast cancer. As described in Example 1, two sets of samples were utilized, which included samples from female individuals having breast cancer (breast cancer cases) and samples from female individuals not having cancer (female controls). The initial screen of each pool was performed in an allelotyping study, in which certain samples in each group were pooled. By pooling DNA from each group, an allele frequency for each SNP in each group was calculated. These allele frequencies were then compared to one another. Particular SNPs were considered as being associated with breast cancer when allele frequency differences calculated between case and control pools were statistically significant. SNP disease association results obtained from the allelotyping study were then validated by genotyping each associated SNP across all samples from each pool. The results of the genotyping were then analyzed, allele frequencies for each group were calculated from the individual genotyping results, and a p-value was calculated to determine whether the case and control groups had statistically significantly differences in allele frequencies for a particular SNP. When the genotyping results agreed with the original allelotyping results, the SNP disease association was considered validated at the genetic level.

SNP Panel Used for Genetic Analyses

[0231] A whole-genome SNP screen began with an initial screen of approximately 25,000 SNPs over each set of disease and control samples using a pooling approach. The pools studied in the screen are described in Example 1. The SNPs analyzed in this study were part of a set of 25,488 SNPs confirmed as being statistically polymorphic as each is characterized as having a minor allele frequency of greater than 10%. The SNPs in the set reside in genes or in close proximity to genes, and many reside in gene exons. Specifically, SNPs in the set are located in exons, introns, and within 5,000 base-pairs upstream of a transcription start site of a gene. In addition, SNPs were selected according to the following criteria: they are located in ESTs; they are located in Locuslink or Ensemble genes; and they are located in Genomatix promoter predictions. SNPs in the set also were selected on the basis of even spacing across the genome, as depicted in Table 2.

[0232] A case-control study design using a whole genome association strategy involving approximately 28,000 single nucleotide polymorphisms (SNPs) was employed. Approximately 25,000 SNPs were evenly spaced in gene-based regions of the human genome with a median inter-marker distance of about 40,000 base pairs. Additionally, approximately 3,000 SNPs causing amino acid substitutions in genes described in the literature as candidates for various diseases were used. The

case-control study samples were of female German origin (German paternal and maternal descent) 548 individuals were equally distributed in two groups (female controls and female cases). The whole genome association approach was first conducted on 2 DNA pools representing the 2 groups. Significant markers were confirmed by individual genotyping.

Table 2

General Statistics		Spacing Statistics	
Total # of SNPs	25,488	Median	37,058 bp
# of Exonic SNPs	>4,335 (17%)	Minimum*	1,000 bp
# SNPs with refSNP ID	20,776 (81%)	Maximum*	3,000,000 bp
Gene Coverage	>10,000	Mean	122,412 bp
Chromosome Coverage	All	Std Deviation	373,325 bp
		**Excludes outliers*	

Allelotyping and Genotyping Results

[0233] The genetic studies summarized above and described in more detail below identified allelic variants associated with breast cancer. The allelic variants identified from the SNP panel described in Table 2 are summarized below in Table 3.

Table 3

SNP Reference	Chromo-some Position	Position in Figs 1-4	Contig Identification	Contig Position	Sequence Identification	Sequence Position	Allelic Variability
rs1949471	198272877	39977	NT_029928	1484976	NM_004087	Exonic (R278Q)	T/C
rs220097	10793860	49860	NT_007819	10345196	NM_014660	intragenic	T/C
rs1990440	71267195	40095	NT_026437	53197195	NM_012074	intragenic	G/C
rs355510	68321769	46769	NT_022778	8587277	NM_001812	intragenic	G/A

[0234] Table 3 includes information pertaining to the incident polymorphic variant associated with breast cancer identified herein. Public information pertaining to the polymorphism and the genomic sequence that includes the polymorphism are indicated. The genomic sequences identified in Table 3 may be accessed at the http address www.ncbi.nih.gov/entrez/query.fcgi, for example, by using the publicly available SNP reference number (*e.g.*, rs1949471). The chromosome position refers to the position of the SNP within NCBI's Genome Build 33, which may be accessed at the following http address: www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=. The "Contig Position" provided in Table 3 corresponds to a nucleotide position set forth in the contig sequence, and designates the polymorphic site corresponding to the SNP reference number. The sequence containing the polymorphisms also may be referenced by the "Sequence Identification" set forth in Table 3. The "Sequence Identification" corresponds to cDNA sequence that encodes associated target polypeptides (*e.g.*, *DLG1*) of the invention. The position of the SNP within the

cDNA sequence is provided in the "Sequence Position" column of Table 3. Also, the allelic variation at the polymorphic site and the allelic variant identified as associated with breast cancer is specified in Table 3. All nucleotide sequences referenced and accessed by the parameters set forth in Table 3 are incorporated herein by reference. rs220097 also is known rs286246.

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0235] A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0236] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer was used to genotype the polymorphism. Table 4 shows PCR primers and Table 5 shows extension primers used for analyzing polymorphisms. The initial PCR amplification reaction was performed in a 5 μl total volume containing 1X PCR buffer with 1.5 mM $MgCl_2$ (Qiagen), 200 μM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

Table 4: PCR Primers

Reference SNP ID	Forward PCR primer	Reverse PCR primer
rs1949471	ACGTTGGATGGCTTCAACTGCTTTGCTATG	ACGTTGGATGTTTCTCAGGGTCAATGACTG
rs220097	GCAAACGTGCACATTTGCAC	TTCCTGGGAATGGATTTCAG
rs1990440	CCAGGGTGTGTTCTAATACG	AAGTCACTAACCCCACACAC
rs355510	TTCTGAGATGATCCTGATGG	CCCTCCTTTTAACCTTTTAGG

[0237] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 μl volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 μl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0238] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™

cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. In Table 5, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

Table 5: Extend Primers

Reference SNP ID	Extend Probe	Term Mix
rs1949471	CAGGGTCAATGACTGTATATTAC	ACT
rs220097	ACAGAGTTTTAAACCTCCTACA	ACT
rs1990440	CGTCAGCAAATGTGTACCGA	ACT
rs355510	ATGGTTTTCTTTCTTGTCCTTC	ACG

[0239] The MassEXTEND™ reaction was performed in a total volume of 9 μl, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0240] Following incubation, samples were desalted by adding 16 μl of water (total reaction volume was 25 μl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP® (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0241] Variations identified in the target genes are provided in their respective genomic sequences (see Figures 1-5) Minor allelic frequencies for these polymorphisms was verified as being 10% or greater by determining the allelic frequencies using the extension assay described above in a group of samples isolated from 92 individuals originating from the state of Utah in the United States, Venezuela and France (Coriell cell repositories).

[0242] Genotyping results are shown for female pools in Table 6A and 6B. Table 6A shows the orginal genotyping results and Table 6B shows the genotyped results re-analyzed to remove duplicate individuals from the cases and controls (*i.e.*, individuals who were erroneously included more than

once as either cases or controls). Therefore, Table 6B represents a more accurate measure of the allele frequencies for this particular SNP. In the subsequent tables, "AF" refers to allelic frequency; and "F case" and "F control" refer to female case and female control groups, respectively.

Table 6A

Reference SNP ID	AF F case	AF F control	p-value	Odds Ratio	Breast Cancer Assoc. Allele
rs1949471	T = 0.186 C = 0.814	T = 0.112 C = 0.890	**0.0005**	0.54	T
rs220097	T = 0.721 C = 0.279	T = 0.626 C = 0.374	**0.0014**	0.66	T
rs1990440	C = 0.876 G = 0.124	C = 0.926 G = 0.074	**0.0027**	0.65	G
rs355510	A = 0.545 G = 0.455	A = 0.616 G = 0.384	**0.0173**	0.75	G

Table 6B

Reference SNP ID	AF F case	AF F control	p-value	Odds Ratio	Breast Cancer Assoc. Allele
rs1949471	T = 0.182 C = 0.818	T = 0.108 C = 0.892	**0.0009**	0.54	T
rs220097	T = 0.709 C = 0.291	T = 0.624 C = 0.376	**0.0045**	0.68	T
rs1990440	C = 0.879 G = 0.121	C = 0.915 G = 0.085	**0.0692**	0.67	G
rs355510	A = 0.539 G = 0.461	A = 0.617 G = 0.383	**0.0123**	0.73	G

[0243] The single marker alleles set forth in Table 3 were considered validated, since the genotyping data for the females, males or both pools were significantly associated with breast cancer, and because the genotyping results agreed with the original allelotyping results. Particularly significant associations with breast cancer are indicated by a calculated p-value of less than 0.05 for genotype results, which are set forth in bold text. Tables 6A and 6B show the disease associated allele in column 6. In the case of rs1949471, this SNP is an exonic SNP that codes for a R278Q amino acid change in the DLG1 gene. The thymine allele codes for glutamine (Q); therefore, a glutamine is associated with an increased risk of breast cancer.

[0244] Odds ratio results are shown in Tables 6A and 6B. An odds ratio is an unbiased estimate of relative risk which can be obtained from most case-control studies. Relative risk (RR) is an estimate of the likelihood of disease in the exposed group (susceptibility allele or genotype carriers) compared to the unexposed group (not carriers). It can be calculated by the following equation:

RR = *I*A/*I*a

*I*A is the incidence of disease in the A carriers and *I*a is the incidence of disease in the non-carriers.

RR > 1 indicates the A allele increases disease susceptibility.

RR < 1 indicates the a allele increases disease susceptibility.

[0245] For example, RR = 1.5 indicates that carriers of the A allele have 1.5 times the risk of disease than non-carriers, *i.e.*, 50% more likely to get the disease.

[0246] Case-control studies do not allow the direct estimation of *I*A and *I*a, therefore relative risk cannot be directly estimated. However, the odds ratio (OR) can be calculated using the following equation:

OR = (nDAnda)/(ndAnDa) = *p*DA(1 – pdA)/*p*dA(1 – pDA), or

OR = ((case f) / (1- case f)) / ((control f) / (1-control f)), where f = susceptibility allele frequency.

[0247] An odds ratio can be interpreted in the same way a relative risk is interpreted and can be directly estimated using the data from case-control studies, *i.e.*, case and control allele frequencies. The higher the odds ratio value, the larger the effect that particular allele has on the development of breast cancer. Possessing an allele associated with a relatively high odds ratio translates to having a higher risk of developing or having breast cancer.

Example 3

DLG1 Region Proximal SNPs

[0248] It has been discovered that a polymorphic variation (rs1949471) in a region that encodes the discs, large homolog 1 (Drosophila) (DLG1) gene is associated with the occurrence of breast cancer (see Examples 1 and 2). Subsequently, SNPs proximal to the incident SNP (rs1949471) were identified and allelotyped in breast cancer sample sets and control sample sets as described in Examples 1 and 2. Approximately twenty-one allelic variants located within the DLG1 region were identified and allelotyped. The polymorphic variants are set forth in Table 7. The chromosome position provided in column four of Table 7 is based on Genome "Build 33" of NCBI's GenBank.

Table 7

dbSNP rs#	Chromosome	Chromosome Position	Position in Figure 1	Allele Variants
2341225	3	198233033	133	T/C
3856760	3	198240838	7938	T/C
2195027	3	198241773	8873	G/A
1356612	3	198246121	13221	C/T
3773845	3	198250188	17288	T/C
2098941	3	198258632	25732	G/A
890491	3	198259823	26923	C/G
1949471	3	198272877	39977	C/T

dbSNP rs#	Chromosome	Chromosome Position	Position in Figure 1	Allele Variants
3773851	3	198274184	41284	T/A
3773852	3	198274310	41410	A/C
3773853	3	198274377	41477	C/T
1195059	3	198274414	41514	G/A
3773855	3	198275506	42606	G/A
3821713	3	198275642	42742	A/C
604005	3	198292415	59515	G/A
DLG1_SNP	3	198292708	59808	T/C
2879969	3	198293165	60265	C/G
958902	3	198300052	67152	T/C
1839742	3	198301232	68332	T/C
1868890	3	198304028	71128	T/C
1868891	3	198309327	76427	G/A

Assay for Verifying and Allelotyping SNPs

[0249] The methods used to verify and allelotype the proximal SNPs of Table 7 are the same methods described in Examples 1 and 2 herein. The PCR primers and extend primers used in these assays are provided in Table 8 and Table 9, respectively.

Table 8

dbSNP rs#	Forward PCR primer	Reverse PCR primer
604005	ACGTTGGATGTGTCTCGCTTTTAGCCTGTG	ACGTTGGATGCAGACAGACATACAGAAGGG
890491	ACGTTGGATGGCAGAACCATGGAGAAAAGC	ACGTTGGATGGGCAAGAGTAAGGCACTATC
958902	ACGTTGGATGGCCACTGAATTGTACATTAAC	ACGTTGGATGATTGGAGTCCCGAGCTAAAC
1195059	ACGTTGGATGCCTGTTTTCATTTAGACTCC	ACGTTGGATGTGCTCACAAAGATTTAAACC
1356612	ACGTTGGATGTTGAACAGCTCAGCTGAAAG	ACGTTGGATGAGATACATGTCTTGTCTGGG
1839742	ACGTTGGATGTCTGAGGTCAGGAGTTTGAG	ACGTTGGATGGCCACCATGTCCAGCTAATT
1868890	ACGTTGGATGAGTGAGGAAGGCCTATTAAC	ACGTTGGATGATACCTGAGTCGAACTCTTG
1868891	ACGTTGGATGTTATTGCTCTTGAACGTGGC	ACGTTGGATGTCTGAGAAAAAGAATTGGGG
1949471	ACGTTGGATGTTTCTCAGGGTCAATGACTG	ACGTTGGATGAGACCCTGCTTCTTTCAACG
2098941	ACGTTGGATGATTAGCTGGGCATGCTATCC	ACGTTGGATGTGTAGCCTTGAATTCCTGGG
2195027	ACGTTGGATGGGCGCTAAATAATGCGCCAC	ACGTTGGATGCTGACCTCGTGATCTGCCTG
2341225	ACGTTGGATGGGCGGGTGGGAAGACTCTAA	ACGTTGGATGTCTTTCACTGTATTCAGATC
2879969	ACGTTGGATGCTCCATTTCAAAAAAAAAAA	ACGTTGGATGCCTTAGAGGTATGTCCAGAG
3773845	ACGTTGGATGACACAAGTAACAAACTTGAG	ACGTTGGATGGTGCTTGAAGAAATTATGTG
3773851	ACGTTGGATGTAAGATACGGAGGATAGAGG	ACGTTGGATGGCATATAGTCTTTGTGGTGTG
3773852	ACGTTGGATGGTGAGTGTACTTAAATAAGTT	ACGTTGGATGGTTTCCCTTTGTGTTTTCAG
3773853	ACGTTGGATGTGGTTTAAATCTTTGTGAGC	ACGTTGGATGCTGTGAGTGTATCTGAAAAC
3773855	ACGTTGGATGGCTTGTTTTATGAACTGGAG	ACGTTGGATGTTAATACCATTGGTTAAATC
3821713	ACGTTGGATGTTCAGGCAACTCAAGTAAGC	ACGTTGGATGTAGAGTGGGTGTTTACACTG
3856760	ACGTTGGATGTGATCTCAGCTCACTGTAAC	ACGTTGGATGTGTAGTCCCAGCTACTCAGG
FCH-1723	ACGTTGGATGGCTTCAACTGCTTTGCTATG	ACGTTGGATGTTTCTCAGGGTCAATGACTG
DLG1_SNP	ACGTTGGATGCTTCATAGTAGCCAGGCTAG	ACGTTGGATGAGCACATGAACAGATGTGTC

Table 9

dbSNP rs#	Extend Primer	Term Mix
604005	TTATCAACCTACAATGGA	ACG
890491	TTATGGCCATACGTAAAAAGCA	ACT
958902	CGGAGGCTTTATTCGTA	ACT
1195059	AAAGATTTAAACCATCAACCAAAT	ACG
1356612	GGGTAGTGGTTTCATGATTTTTA	ACG
1839742	TCCAGCTAATTTTTTGTATTTTTA	ACT
1868890	CTGAGTCGAACTCTTGTATAAA	ACT
1868891	GAAAAAGAATTGGGGATTATAAC	ACG
1949471	CGAACATCTACTTCATTTACT	ACG
2098941	TCCTCCCACATCAGCCT	ACG
2195027	GCGTGAGCCACCACACC	ACG
2341225	CACTGTATTCAGATCTTCATATTT	ACT
2879969	CATCATACTGCCTCTGG	ACT
3773845	TTATGTGTTCTCTATTTATTGACT	ACT
3773851	TTTGTGGTGTGGGATTC	CGT
3773852	TATTTTCCATTTCCTCTCTG	ACT
3773853	AAGGGAAACTCATGATTTCTA	ACG
3773855	AGGCTTTTTGTAGCAGT	ACG
3821713	GTGGGTGTTTACACTGTTTAATAC	ACT
3856760	ATGAGAATCACTTGAACCTG	ACT
FCH-1723	CAGGGTCAATGACTGTATATTAC	ACT
DLG1_SNP	AGATGTGTCACAAATGCAA	ACT

Genetic Analysis of Allelotyping Results

[0250] Allelotyping results are shown for cases and controls in Table 10. The allele frequency for the A2 allele is noted in the fifth and sixth columns for breast cancer pools and control pools, respectively, where "AF" is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs2341225 has the following case and control allele frequencies: case A1 (T) = 0.747; case A2 (C) = 0.253; control A1 (T) = 0.743; and control A2 (C) = 0.257, where the nucleotide is provided in paranthesis. SNPs with blank allele frequencies were untyped.

Table 10

dbSNP rs#	Chromosome	Position in Figure 1	Allele Variants	A2 Case AF	A2 Control AF	p-Value
2341225	198233033	133	T/C	0.253	0.257	0.8897
3856760	198240838	7938	T/C	0.959	0.985	0.0095
2195027	198241773	8873	G/A	0.651	0.691	0.1538
1356612	198246121	13221	C/T	0.197	0.243	0.0653
3773845	198250188	17288	T/C	0.415	0.414	0.9646
2098941	198258632	25732	G/A	0.281	0.335	0.0515
890491	198259823	26923	C/G	0.440	0.525	0.0051

dbSNP rs#	Chromosome	Position in Figure 1	Allele Variants	A2 Case AF	A2 Control AF	p-Value
1949471	198272877	39977	C/T	0.181	0.092	0.0001
3773851	198274184	41284	T/A	0.351	0.371	0.4824
3773852	198274310	41410	A/C	0.206	0.233	0.2786
3773853	198274377	41477	C/T	0.485	0.480	0.8660
1195059	198274414	41514	G/A	0.936	0.931	0.7361
3773855	198275506	42606	G/A	0.275	0.260	0.5723
3821713	198275642	42742	A/C	0.728	0.677	0.0666
604005	198292415	59515	G/A	0.985	0.986	0.8647
DLG1_SNP	198292708	59808	T/C	0.723	0.825	0.0002
2879969	198293165	60265	C/G	0.589	0.596	0.8093
958902	198300052	67152	T/C	0.215	0.264	0.0568
1839742	198301232	68332	T/C	0.928	0.946	0.2311
1868890	198304028	71128	T/C	0.420	0.422	0.9494
1868891	198309327	76427	G/A	0.220	0.217	0.8858

[0251] Figure 13 shows the proximal SNPs in and around the DLG1 gene. The position of each SNP on the chromosome is presented on the x-axis. The y-axis gives the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 13 can be determined by consulting Table 10. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0252] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). The black line (or generally top-most curve, *e.g.*, see peak in left-most graph just to the left of position 92150000) provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0253] Finally, the gene or genes present in the loci region of the proximal SNPs as annotated by Locus Link (http address: www.ncbi.nlm.nih.gov/LocusLink/) are provided on the graph. The exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon

positions are shown as thick, unbroken bars. An arrow is place at the 3' end of each gene to show the direction of transcription.

Example 4

KIAA0783 Proximal SNPs

[0254] It has been discovered that a polymorphic variation (rs220097) in a region that encodes KIAA0783 is associated with the occurrence of breast cancer (see Examples 1 and 2). Subsequently, SNPs proximal to the incident SNP (rs220097) were identified and allelotyped in breast cancer sample sets and control sample sets as described in Examples 1 and 2. Approximately fifty-eight allelic variants located within the KIAA0783 region were identified and allelotyped. The polymorphic variants are set forth in Table 11.

Table 11

dbSNP rs#	Chromosome	Position in Figure 2	Chromosome Position	Allele Variants
218973	7	201	10710201	G/A
218962	7	6395	10716395	T/C
1640705	7	8558	10718558	T/C
218983	7	9429	10719429	C/T
190075	7	9809	10719809	T/G
284856	7	10072	10720072	C/T
218981	7	10511	10720511	C/T
218980	7	11556	10721556	C/G
1640703	7	16857	10726857	A/G
1640702	7	16951	10726951	A/G
1640701	7	17027	10727027	C/G
1681305	7	17177	10727177	T/C
1640700	7	17615	10727615	A/C
1640699	7	17950	10727950	C/G
1154923	7	18329	10728329	T/G
1154922	7	18384	10728384	T/C
1154921	7	18561	10728561	G/A
1154920	7	18579	10728579	C/T
2510348	7	18871	10728871	C/G
1681311	7	27152	10737152	C/T
1681312	7	27306	10737306	T/C
1681286	7	28091	10738091	T/C
1640710	7	28661	10738661	A/C
1681284	7	29011	10739011	T/C
2110377	7	29962	10739962	T/G
2110376	7	29969	10739969	T/G
2160059	7	30085	10740085	T/C
1681290	7	31656	10741656	A/G
1681291	7	31685	10741685	A/G
1681292	7	31749	10741749	G/A
220091	7	45389	10755389	T/C
182594	7	45459	10755459	G/C
220090	7	46647	10756647	A/G

dbSNP rs#	Chromosome	Position in Figure 2	Chromosome Position	Allele Variants
220097	7	49860	10759860	T/C
220096	7	53061	10763061	T/C
220095	7	57308	10767308	T/A
3801435	7	61563	10771563	A/G
1681281	7	61660	10771660	A/G
1026903	7	62212	10772212	C/T
220093	7	67090	10777090	T/G
286243	7	67198	10777198	T/C
3801437	7	70071	10780071	A/G
3801438	7	70191	10780191	G/A
2108111	7	74006	10784006	C/T
2353340	7	75600	10785600	A/G
3823875	7	85761	10795761	A/G
2190295	7	90798	10800798	T/G
KIAA0783_SNP1	7	90883	10800883	C/T
2306768	7	91259	10801259	T/A
2353341	7	95416	10805416	C/G
2353342	7	95446	10805446	T/G
2883140	7	96368	10806368	G/T
2353343	7	97050	10807050	T/C
2108114	7	97362	10807362	C/T
1483204	7	97630	10807630	A/C
1483202	7	97989	10807989	T/C
1483201	7	98107	10808107	C/T
KIAA0783_SNP2	7	NOT MAPPED		

Assay for Verifying and Allelotyping SNPs

[0255] The methods used to verify and allelotype the proximal SNPs of Table 11 are the same methods described in Examples 1 and 2 herein. The PCR primers and extend primers used in these assays are provided in Table 12 and Table 13, respectively.

Table 12

dbSNP rs#	Forward PCR primer	Reverse PCR primer
KIAA0783_SNP1	ACGTTGGATGCCCTAACACTACTCCTTGTC	ACGTTGGATGCCAACACTTCTTGGAGTCTG
KIAA0783_SNP2	ACGTTGGATGAGCCACATTCTCAGATACTG	ACGTTGGATGGGAAAAGAAGGAAGAAGAAG
182594	ACGTTGGATGGAGACAGAAAAGTGGTGGAC	ACGTTGGATGCCTTTAAGAAGGCCCTTGTG
190075	ACGTTGGATGCACAAATTCAGTGGCCAAGC	ACGTTGGATGCTTGTTGTGGACACCTACTG
218962	ACGTTGGATGCAGGAGTGAGAAGTTCTTTG	ACGTTGGATGTGCTGATTGGTCTATGGGTG
218973	ACGTTGGATGTCTCACACTGAGGCCTGTAG	ACGTTGGATGTTTGCTGCACCCATCAACTC
218980	ACGTTGGATGCTTCCCTCCTTTTCTCCTTC	ACGTTGGATGCAAGATCCAATCCAGAAGAC
218981	ACGTTGGATGAGATTGCTGCCACTACACAC	ACGTTGGATGCTCTTGGCATTCTTAACTCAG
218983	ACGTTGGATGTCTGCAGTTTCTCTCTCAAC	ACGTTGGATGACCAAATCCAAGATGTAGGG
220090	ACGTTGGATGCAGCAGAAACTTGATGATGG	ACGTTGGATGAGACACTGAGACTCTGGAGG

dbSNP rs#	Forward PCR primer	Reverse PCR primer
220091	ACGTTGGATGGTGTATACACAAGGGCCTTC	ACGTTGGATGCTGATTGCTGTTTCTGTTAC
220093	ACGTTGGATGTCCACACTGTGAACAGAGAC	ACGTTGGATGAGTCTAAAAAGGCTGTCAGG
220095	ACGTTGGATGGCAGCTCAATTTTTAGGAACC	ACGTTGGATGCCCTTGTACACTGTTGCATG
220096	ACGTTGGATGTAGATTAATTATTGGTTGGC	ACGTTGGATGGCCACCTCCAAAATTAGATC
220097	ACGTTGGATGTTCCTGGGAATGGATTTCAG	ACGTTGGATGGCAAACGTGCACATTTGCAC
284856	ACGTTGGATGTGCATGACTACACAAAGAAG	ACGTTGGATGGCAAAATCCTACATTGAGGC
286243	ACGTTGGATGATGTCTCTGTTCACAGTGTG	ACGTTGGATGCTGGCAAATAGCAATCTAAAC
220097	ACGTTGGATGTTCCTGGGAATGGATTTCAG	ACGTTGGATGGCAAACGTGCACATTTGCAC
1026903	ACGTTGGATGGTACTGAAACTCTGAGCATTC	ACGTTGGATGCATCTTATCTGTTTACCATAC
1154920	ACGTTGGATGGCTGTATATACGAGTTAATGG	ACGTTGGATGAGTGGAGGTGGAGGTGAGGCT
1154921	ACGTTGGATGAAATGCCAATAGCGCCAAGG	ACGTTGGATGAGTAGAAGAGATAAGCCTGG
1154922	ACGTTGGATGTTTTGCCTCACCAAGATTGC	ACGTTGGATGACAATTTCATTGAGGAGAGG
1154923	ACGTTGGATGGATGGTTGATCACTTGTGTAG	ACGTTGGATGCTTACCTCCTCTCCTCAATG
1483201	ACGTTGGATGGTTGCTAAAGTAGTTTCAGCC	ACGTTGGATGACCAAAGAGCTTGTCCCATC
1483202	ACGTTGGATGGTGCTTTAGAATGTAACACAG	ACGTTGGATGTGGAATTGCACCCTTGCTTG
1483204	ACGTTGGATGTATCTTATCTAGCAGGCAAC	ACGTTGGATGACTAAGATCACAGGCCTGAG
1640699	ACGTTGGATGGGTTGGGTGTATGATAGGAG	ACGTTGGATGAGCATGGCTAATCTGTCTGG
1640700	ACGTTGGATGCTTTATTGACTGCTTTCAATC	ACGTTGGATGAGTGATTACGAGCCTGTACC
1640701	ACGTTGGATGTTAGGTGCATTGATGCTCTG	ACGTTGGATGCTCAGGCACAGAAAAGATTC
1640702	ACGTTGGATGCTGTGGTCTCAGGTCACAAA	ACGTTGGATGATGCACCTAAAACAAGAGTC
1640703	ACGTTGGATGCATAATTTACCTTCCTGGCC	ACGTTGGATGCAAATTTGTGACCTGAGACC
1640705	ACGTTGGATGACCATCAGAACCAGTATACC	ACGTTGGATGGATGGCCAGAATTGATGTAC
1640710	ACGTTGGATGCCTTTCCGCTGTATCTCTTG	ACGTTGGATGGGTACAAGGAAGATCCTCAG
1681281	ACGTTGGATGATTGAGAAAGCAGCTGCTTG	ACGTTGGATGCCAACCTCCCAAATACATCC
1681284	ACGTTGGATGATAAAATAGGTCTGGGGCTG	ACGTTGGATGGTTTGCTTACTCTGGTACTG
1681286	ACGTTGGATGGAAATGTAACGCAAAGAGGG	ACGTTGGATGGTTGAAACATTGTCTGCTAG
1681290	ACGTTGGATGGTACCATAAAATACAATACC	ACGTTGGATGTGGTCCCCCAGTCATCTTAA
1681291	ACGTTGGATGTAGCAAAACCCTGCCTCTAC	ACGTTGGATGAGGTCAGTGTTCTGGTATTG
1681292	ACGTTGGATGAGGTCAGTGTTCTGGTATTG	ACGTTGGATGAGCCTGGGCAACATAGCAAA
1681305	ACGTTGGATGCAGACAGATGTTTAGCTACC	ACGTTGGATGTGAAGTTGTGGATTCCCAGC
1681311	ACGTTGGATGGCTTGACCAATCATACTTCC	ACGTTGGATGGAAACAAATTGCTCTGAGTCC
1681312	ACGTTGGATGTCTTCAGGGCAGTAGGATTC	ACGTTGGATGCACATGTGTTTAATACAAGG
2108111	ACGTTGGATGAGCCTGTAAATGATAGAGCC	ACGTTGGATGGATGTCACAGTACAGCAATG
2108114	ACGTTGGATGGATAGAAAAGTTAGAGAAATG	ACGTTGGATGAAGGTCACACCACTGCACTC
2110376	ACGTTGGATGCCAGTTTACACTGGATATTTC	ACGTTGGATGTTGACTAGCTGCTAGAAGGG
2110377	ACGTTGGATGCCAGTTTACACTGGATATTTC	ACGTTGGATGTTGACTAGCTGCTAGAAGGG
2160059	ACGTTGGATGTTAAGTACCGGGAAATTCAG	ACGTTGGATGTCATATACCTACGCAGGCTC
2190295	ACGTTGGATGCTTTTAGAAGTAGTAGGGGC	ACGTTGGATGAGACTCCAAGAAGTGTTGGG
2306768	ACGTTGGATGAAAGGTGGTTTTGCCAGCTG	ACGTTGGATGCTCAGTCTCCTGAAGTGCTG
2353340	ACGTTGGATGCCTATCTGCATGTTGCTTAC	ACGTTGGATGGACTCTTGGGAGTACAAATG
2353341	ACGTTGGATGCACAACCAGAATTTGTAAGTC	ACGTTGGATGCACACGCATGCATCATCTAC
2353342	ACGTTGGATGTGGTTTTCAGTCAAAGCTGC	ACGTTGGATGCTGAGATCTTTCTTCCTGAC
2353343	ACGTTGGATGGTTGCAGAGGGAAGCATTTC	ACGTTGGATGCACTTGTGACCAGGTCACTA
2510348	ACGTTGGATGCTATCCCAGGGCTATGTTTG	ACGTTGGATGGAAGTGGAGGATGAGTTGTG
2883140	ACGTTGGATGCAGCACTTACTTGTCATGTAG	ACGTTGGATGCATAACCAATTTGTCTTAAC
3801435	ACGTTGGATGTCAGTATGAAGCAAGCAGCC	ACGTTGGATGATGTCGCTATACTCTGTAGG
3801437	ACGTTGGATGGTAGCTGAGAAGATGCTCAC	ACGTTGGATGATAGCTGTTCCAGTCTCTTG
3801438	ACGTTGGATGATACGGTAAAGGTAGTCTGG	ACGTTGGATGTTACCTGTATTGCCCTCTCG
3823875	ACGTTGGATGCTCAAGAGCCCATCATCATC	ACGTTGGATGGACAGGCTCAGATATTTCAG

Table 13

dbSNP rs#	Extend Primer	Term Mix
KIAA0783_SNP1	ATTCAGCACAAGTTGTCA	ACG
KIAA0783_SNP2	GAAAGACCTAGAAAGAAAA	ACT
182594	CTCTCTCTTTCTCTCACT	ACT
190075	GTCTGGAGATCCGAATTT	ACT
218962	GCACCATCTGATTGGCC	ACT
218973	CCCAACACTATCCCTTC	ACG
218980	ATCCAGAAGACAATATTGCATTTA	ACT
218981	GTATTGCTTTGTTGCCC	ACG
218983	GGTAAAGAGATGAAGTGC	ACG
220090	CCCAGATATCCTCGGAA	ACT
220091	TGTTACTTATTACATTGTCCAA	ACT
220093	TTATATTCACTCTGAAATCCC	ACT
220095	CACTGTTGCATGAAATGTA	CGT
220096	CCTGCTACAAAGGGACCTCA	ACT
220097	ACAGAGTTTTAAACCTCCTACA	ACT
284856	TACATTGAGGCAGTTTGTGCT	ACG
286243	AGCAATCTAAACATGAGATTGAGC	ACT
220097	ACAGAGTTTTAAACCTCCTACA	ACT
1026903	CTTATCTGTTTACCATACAATCTA	ACG
1154920	CAACACAAAATGCCAATAG	ACG
1154921	TGTGGCTGTATATACGAGTTAA	ACG
1154922	TTGAGGAGAGGAGGTAA	ACT
1154923	CATCAATCTAATCTCATTTCCTAT	ACT
1483201	TGGGTGGTCCTTTTCTGATA	ACG
1483202	TAATCATGTGGAATTTCCAG	ACT
1483204	CAGGCCTGAGCCACTGT	ACT
1640699	CTAATCTGTCTGGTTAATAGAA	ACT
1640700	GCAAAAGCAAAAGTAAGCT	ACT
1640701	AAACAATGGTAATCTAGAGTAAGC	ACT
1640702	TGATTCAATTTCTGTTGACTACT	ACT
1640703	GTGACCTGAGACCACAGATC	ACT
1640705	TCCAAATAAGAAGCCCT	ACT
1640710	CAGTGTAATAAATTATCAGTTCAT	ACT
1681281	TGGAGTTCAATATAAAGATACAC	ACT
1681284	TGTTTTCAGTTTTATTTGCC	ACT
1681286	TTGTCTGCTAGCCATTT	ACT
1681290	AATCAGTGTTTCTTTAAAGGTC	ACT
1681291	CTGGTATTGTATTTTATGGTACT	ACT
1681292	GGGCAACATAGCAAAACCCTG	ACG
1681305	TTCCCAGCCCTACTTAC	ACT
1681311	CTGAGTCCTAAAAAAAGGT	ACG
1681312	TTAATACAAGGAAATTCCAGC	ACT
2108111	AGAATTTGAAGACATAAAAACC	ACG
2108114	GCGACAGAGCAAGACTC	ACG
2110376	GGGTCAGAGAACTCTATTAA	ACT
2110377	AGAGAACTCTATTAAGTAGGTC	ACT

dbSNP rs#	Extend Primer	Term Mix
2160059	CTCATGGATCTGTCTTAC	ACT
2190295	GGGGAAAAAAAAGGTCATATTA	ACT
2306768	CTGAAGTGCTGGGATTATGGG	CGT
2353340	TTTTCTGTGCTTTCTTTGT	ACT
2353341	CATCTACTCTCTTTGAAGTT	ACT
2353342	CTTTCTTCCTGACTTACAAATTC	ACT
2353343	GTGTTTTTGTTGACATATCAAT	ACT
2510348	GGAGGATGAGTTGTGTTGACT	ACT
2883140	TTGTCTTAACTACTATAAACTGAA	CGT
3801435	GCTATACTCTGTAGGAGTTTATCT	ACG
3801437	CAGTCTCTTGATTTTAAGGA	ACT
3801438	CTCGTACTTTTGCCCAC	ACG
3823875	ATTTCAGTGATATAGGAGTCT	ACT

Genetic Analysis of Allelotyping Results

[0256] Allelotyping results are shown for cases and controls in Table 14. The allele frequency for the A2 allele is noted in the fifth and sixth columns for breast cancer pools and control pools, respectively, where "AF" is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs218973 has the following case and control allele frequencies: case A1 (G) = 0.640; case A2 (A) = 0.360; control A1 (G) = 0.645; and control A2 (A) = 0.355, where the nucleotide is provided in paranthesis. SNPs with blank allele frequencies were untyped.

Table 14

dbSNP rs#	Position in Figure 2	Chromosome Position	A1/A2 Allele	A2 Case AF	A2 Control AF	p-Value
218973	201	10710201	G/A	0.360	0.355	0.8462
218962	6395	10716395	T/C	0.547	0.535	0.6939
1640705	8558	10718558	T/C	0.601	0.568	0.2583
218983	9429	10719429	C/T	0.561	0.558	0.9406
190075	9809	10719809	T/G	0.447	0.428	0.5348
284856	10072	10720072	C/T	0.612	0.585	0.3555
218981	10511	10720511	C/T	0.432	0.363	0.0189
218980	11556	10721556	C/G	0.409	0.471	0.0378
1640703	16857	10726857	A/G	0.841	0.859	0.3809
1640702	16951	10726951	A/G	0.674	0.656	0.5269
1640701	17027	10727027	C/G	0.266	0.270	0.9020
1681305	17177	10727177	T/C	0.422	0.483	0.0406
1640700	17615	10727615	A/C	0.456	0.423	0.2641
1640699	17950	10727950	C/G	0.344	0.370	0.3558
1154923	18329	10728329	T/G	0.885	0.878	0.7144
1154922	18384	10728384	T/C	0.406	0.479	0.0151
1154921	18561	10728561	G/A	0.367	0.365	0.9611
1154920	18579	10728579	C/T	0.284	0.248	0.1803
2510348	18871	10728871	C/G	0.409	0.425	0.5940
1681311	27152	10737152	C/T	0.251	0.279	0.3099
1681312	27306	10737306	T/C	0.303	0.260	0.1171
1681286	28091	10738091	T/C	0.557	0.544	0.6560

dbSNP rs#	Position in Figure 2	Chromosome Position	A1/A2 Allele	A2 Case AF	A2 Control AF	p-Value
1640710	28661	10738661	A/C	0.455	0.515	0.0472
1681284	29011	10739011	T/C	0.418	0.388	0.3124
2110377	29962	10739962	T/G	0.080	0.058	0.1549
2110376	29969	10739969	T/G	0.265	0.313	0.0798
2160059	30085	10740085	T/C	0.066	0.063	0.8793
1681290	31656	10741656	A/G	0.222	0.287	0.0129
1681291	31685	10741685	A/G	0.017	0.042	0.0143
1681292	31749	10741749	G/A	0.335	0.392	0.0458
220091	45389	10755389	T/C	0.245	0.326	0.0034
182594	45459	10755459	G/C	0.238	0.325	0.0017
220090	46647	10756647	A/G	0.332	0.411	0.0066
220097	49860	10759860	T/C	0.258	0.343	0.0025
220096	53061	10763061	T/C	0.240	0.301	0.0214
220095	57308	10767308	T/A	0.618	0.526	0.0026
3801435	61563	10771563	A/G	0.622	0.507	0.0002
1681281	61660	10771660	A/G	0.501	0.433	0.0235
1026903	62212	10772212	C/T	0.855	0.859	0.8503
220093	67090	10777090	T/G	0.564	0.461	0.0009
286243	67198	10777198	T/C	0.591	0.519	0.0170
3801437	70071	10780071	A/G	0.385	0.459	0.0141
3801438	70191	10780191	G/A	0.018	0.022	0.6491
2108111	74006	10784006	C/T	0.360	0.438	0.0090
2353340	75600	10785600	A/G	0.234	0.309	0.0056
3823875	85761	10795761	A/G	0.502	0.409	0.0025
2190295	90798	10800798	T/G	0.319	0.402	0.0045
KIAA0783_SNP1	90883	10800883	C/T	0.309	0.396	0.0030
2306768	91259	10801259	T/A	0.558	0.472	0.0051
2353341	95416	10805416	C/G	0.163	0.248	0.0008
2353342	95446	10805446	T/G	0.118	0.176	0.0068
2883140	96368	10806368	G/T	0.672	0.561	0.0003
2353343	97050	10807050	T/C	0.071	0.075	0.8073
2108114	97362	10807362	C/T	0.433	0.321	0.0003
1483204	97630	10807630	A/C	0.063	0.093	0.0706
1483202	97989	10807989	T/C	0.643	0.567	0.0101
1483201	98107	10808107	C/T	0.688	0.598	0.0022
KIAA0783_SNP2	NOT MAPPED			0.411	0.459	0.1085

[0257] Figure 14 shows the proximal SNPs in and around the KIAA0783 region. The position of each SNP on the chromosome is presented on the x-axis. The y-axis gives the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 14 can be determined by consulting Table 14. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0258] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie,

Wadsworth & Brooks/Cole.). The black line (or generally top-most curve, *e.g.*, see peak in left-most graph just to the left of position 92150000) provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0259] Finally, the gene or genes present in the loci region of the proximal SNPs as annotated by Locus Link (http address: www.ncbi.nlm.nih.gov/LocusLink/) are provided on the graph. The exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is place at the 3' end of each gene to show the direction of transcription.

Example 5

DPF3 Proximal SNPs

[0260] It has been discovered that a polymorphic variation (rs1990440) in a gene encoding DPF3 is associated with the occurrence of breast cancer (see Examples 1 and 2). Subsequently, SNPs proximal to the incident SNP (rs1990440) were identified and allelotyped in breast cancer sample sets and control sample sets as described in Examples 1 and 2. A total of forty allelic variants located within or nearby the DPF3 gene were identified and allelotyped. The polymorphic variants are set forth in Table 15. The chromosome position provided in column four of Table 15 is based on Genome "Build 33" of NCBI's GenBank.

Table 15

dbSNP rs#	Chromosome	Position in Figure 3	Chromosome Position	Allele Variants
2052146	14	160	71227260	A/C
2052145	14	6053	71233153	T/G
740980	14	9719	71236819	A/G
758915	14	10481	71237581	T/C
758914	14	10676	71237776	A/T
2098195	14	17179	71244279	C/G
740979	14	18561	71245661	A/T
740978	14	18658	71245758	G/C
740977	14	18694	71245794	A/G
740976	14	18858	71245958	T/C
2052143	14	24582	71251682	G/A
2052142	14	24683	71251783	G/A
2052141	14	24767	71251867	C/T
758913	14	27402	71254502	A/G
740975	14	28150	71255250	T/G
747987	14	28494	71255594	T/C

dbSNP rs#	Chromosome	Position in Figure 3	Chromosome Position	Allele Variants
1126160	14	32003	71259103	A/C
2332918	14	35588	71262688	C/T
2332919	14	35619	71262719	T/C
1990443	14	35856	71262956	G/A
3937455	14	36254	71263354	G/C
973963	14	37314	71264414	G/A
1990441	14	40033	71267133	T/G
1990440	14	40095	71267195	G/C
2159715	14	42593	71269693	A/C
2109795	14	42799	71269899	A/G
2159714	14	43090	71270190	G/A
1468662	14	46683	71273783	A/G
2215591	14	49774	71276874	A/G
2109794	14	51796	71278896	C/T
2877821	14	52079	71279179	A/T
2191822	14	53857	71280957	T/C
2191821	14	53971	71281071	A/C
1544579	14	55899	71282999	T/C
2215590	14	60682	71287782	G/A
1004552	14	61291	71288391	C/T
1860749	14	72720	71299820	G/A
1860748	14	72752	71299852	A/C
763388	14	85507	71312607	A/G
1035099	14	89751	71316851	T/A

Assay for Verifying and Allelotyping SNPs

[0261] The methods used to verify and allelotype the sixty-three proximal SNPs of Table 15 are the same methods described in Examples 1 and 2 herein. The PCR primers and extend primers used in these assays are provided in Table 16 and Table 17, respectively.

Table 16

dbSNP rs#	Forward PCR primer	Reverse PCR primer
740975	ACGTTGGATGGAAACCAAGATAGGAAATGG	ACGTTGGATGCTCAGTGCCAGAAATACCAG
740976	ACGTTGGATGTCCTGTTTCTAAGCAGGGAG	ACGTTGGATGATCAGGACTACCTGAGCAAC
740977	ACGTTGGATGTCCAGTGAGGCCTCCCTCCAA	ACGTTGGATGCAGCAACCCAAAGCAACACG
740978	ACGTTGGATGTAGCCACGCCATTATTGGAG	ACGTTGGATGCTTCACATCCCTCCTCAAAG
740979	ACGTTGGATGATCCTAACCAGGTCTGATGG	ACGTTGGATGAAGGGCCAAGCAATGCTTTG
740980	ACGTTGGATGGGTAGGGCTGTCTGTTTCAT	ACGTTGGATGATGCCTGCCACATTGGGTAA
747987	ACGTTGGATGAGGTCTGGCACTGCTAAATG	ACGTTGGATGCCTTGTGAACTTCCAACCTG
758913	ACGTTGGATGCCTAGCCAACATCCTTTTCC	ACGTTGGATGAGCAACCAGTCTAGTTTTCG
758914	ACGTTGGATGCCCTTGTTTTAGAGGTTGGG	ACGTTGGATGTGTGATCCAGACATCAGCTC
758915	ACGTTGGATGCAAGAAGGGCATTTCTACCC	ACGTTGGATGCAATGCTGCTGACATCAGAC
763388	ACGTTGGATGGGGTACTCTTAGCTGAGAAC	ACGTTGGATGTACAGGGATTGTGATGTGGG
973963	ACGTTGGATGGATTTGTTCTGGCAGGAATG	ACGTTGGATGACAAACCACTAAACTTTCAG
1004552	ACGTTGGATGGATCATCCAAGTATGCTCCC	ACGTTGGATGGCAAAACCCAGTGCCAAAAC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
1035099	ACGTTGGATGAAAGGGTACCCAGACTTCAC	ACGTTGGATGTGGGGAGAACTTTGGTCAAC
1126160	ACGTTGGATGGGGTTCTCTCTTGACAGATG	ACGTTGGATGTGTTCTCACCCTGTTCTGTT
1468662	ACGTTGGATGGCTAGAAATCACCAGCAACC	ACGTTGGATGTCATGTAGGTTGGCTCTGAC
1544579	ACGTTGGATGACCATTATCATCTTCCCAGG	ACGTTGGATGCCTTATCTCTCTAAGACATGC
1860748	ACGTTGGATGACTCGACTAGCTAGTCTTGG	ACGTTGGATGAAAGCAATCCAGCGGACAAG
1860749	ACGTTGGATGTCCCCGGAATGATACATGAC	ACGTTGGATGAACATGATTAAGGATAAAGC
1990440	ACGTTGGATGAAGTCACTAACCCCACACAC	ACGTTGGATGCCAGGGTGTGTTCTAATACG
1990441	ACGTTGGATGTCAGAGATATGCACTGCAAG	ACGTTGGATGCACACCCTGGCATGAATGTG
1990443	ACGTTGGATGCACTGGATTTGGCAAGAAGG	ACGTTGGATGTACATGATCCTCCCCTCTAC
2052141	ACGTTGGATGCCTGCAAAATCCCTCATACC	ACGTTGGATGATAGAAGCGTGACCTTACCC
2052142	ACGTTGGATGGGTATGAGGGATTTTGCAGG	ACGTTGGATGACTGGACTCACCCACATAAG
2052143	ACGTTGGATGCCAGTGTAATCACAAGGGTC	ACGTTGGATGTGTGTCACTTCTACCTCCAC
2052145	ACGTTGGATGGTGCTGGCTGCCTAGTTCTA	ACGTTGGATGGGCTTCTCAATTCAGATGGG
2052146	ACGTTGGATGCCACAAAAGCACGTGATTTC	ACGTTGGATGTTATTTGAGCTCTGATAGTG
2098195	ACGTTGGATGGCTCCAGTCTCTAATCACAC	ACGTTGGATGCAAAGTTCTCTGCCTGAGTG
2109794	ACGTTGGATGTAATCCCAGCACTTTGGGAG	ACGTTGGATGAGGCTGATCTTGAACTCCTG
2109795	ACGTTGGATGCAAACAAGGTCCCAGCATTC	ACGTTGGATGTCCTGACTCTCTCAAAACCC
2159714	ACGTTGGATGAAACTCTCTCGTTGCTGTGG	ACGTTGGATGAAAGCCCCTCTAGCAAAAGG
2159715	ACGTTGGATGCTGCCTGCAAGTTCCCATTG	ACGTTGGATGTACAGGCACTGGCGAAGAAG
2191821	ACGTTGGATGGAAAGTGTCCTTAGCTTGCC	ACGTTGGATGTGAGATGGATCTGGAGCCAC
2191822	ACGTTGGATGATTTTTCCCGGCATCTGACC	ACGTTGGATGTGCAAAGTGGTGGAGGAAAG
2215590	ACGTTGGATGTCCAAGAAGGACAGCAGTAG	ACGTTGGATGATGAGAGCCTTTCTTCAGGG
2215591	ACGTTGGATGATTTGTTAAAATTCATAGAAC	ACGTTGGATGTCCCCAGTTTGCATCTTGAC
2332918	ACGTTGGATGAACCCATGGGACCACAATTC	ACGTTGGATGTAGGATGGGTGTTTCCTAGC
2332919	ACGTTGGATGTCTGAGGGCTCTCTCTAATG	ACGTTGGATGATGAAGGAAGAAGCCCTGAC
2877821	ACGTTGGATGATAATCTATGTCCTAGATTG	ACGTTGGATGTAGTAGCATTCCAAGTGCCC
3937455	ACGTTGGATGGCAAGAATAGGTTCTTTCGC	ACGTTGGATGACCTCCACACTCATTACCTC

Table 17

dbSNP rs#	Extend Primer	Term Mix
740975	ACCAGCTCTCTTTGGAT	ACT
740976	ATCCAGATGGCCCTGAC	ACT
740977	TGGTTTTCGAATAAGTAGCCAC	ACT
740978	AAGCCTTCCTATCCCCA	ACT
740979	TGCTTTGGGGCAGACTGAC	CGT
740980	CACATTGGGTAAATGATGA	ACT
747987	AACCTGGTTCTGCCATT	ACT
758913	CCAGTCTAGTTTTCGATCACC	ACT
758914	CCCCAGTGATCCTGAGAAAT	CGT
758915	GACATCAGACCTATGCCAGGA	ACT
763388	CACTCATGCCTCAAGCCAAT	ACT
973963	AACAACCAACTCTCCAG	ACG
1004552	TCTTGGCTCAGTGCTGC	ACG
1035099	TTGGTCAACATCGCAGC	CGT
1126160	GAAGCCCATCGCTAAGTGTTT	ACT

dbSNP rs#	Extend Primer	Term Mix
1468662	CTCTGACTGAGGAGAGACC	ACT
1544579	GACATGCATCAAAGCAGCTG	ACT
1860748	TCTTGGAGCCATATTTTATTTG	ACT
1860749	TTAAGGATAAAGCAATCCAG	ACG
1990440	CGTCAGCAAATGTGTACCGA	ACT
1990441	CATGAATGTGATTCACATTCTCC	ACT
1990443	TTCCCCTCAGCTCTTAG	ACG
2052141	CTTACCCCCAAAGATGTCCA	ACG
2052142	AGCCAGGATAATCTCCTCA	ACG
2052143	TCTACCTCCACTTCCAA	ACG
2052145	ATTCAGATGGGATCACAGAAG	ACT
2052146	GAGCTCTGATAGTGATTGTGAGT	ACT
2098195	TAAACCTTTCTATGTTCCTG	ACT
2109794	CTCAGGTGATCCACCCA	ACG
2109795	TCCCAGAATTTGGAGCC	ACT
2159714	CAAAAGGATCTGCAAAAG	ACG
2159715	CATAGGGATAGGAATGGG	ACT
2191821	ATGTGGGTTTGGACTGGGGCT	ACT
2191822	AGGAAAGGAATGTCTGCCCC	ACT
2215590	CAGGGCCAGCCATGAACGT	ACG
2215591	TTCAATAAAATGTACTCATTCAAA	ACT
2332918	TCTCTCTAATGGGGACC	ACG
2332919	ACTGGATCCCAGAAGAG	ACT
2877821	CCCTGTTCTGCACCTTTAAA	CGT
3937455	TCCTTTTTTCCCCACCC	ACT

Genetic Analysis of Allelotyping Results

[0262] Allelotyping results are shown for cases and controls in Table 18. The allele frequency for the A2 allele is noted in the fifth and sixth columns for breast cancer pools and control pools, respectively, where "AF" is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP in row 2 of Table 13 (rs2052146) has the following case and control allele frequencies: case A1 (A) = 0.990; case A2 (C) = 0.010; control A1 (A) = 0.948; and control A2 (C) = 0.052, where the nucleotide is provided in parenthesis. SNPs with blank allele frequencies were untyped ("not AT").

Table 18

dbSNP rs#	Position in Fig 3	Chrom Position	Alleles (A1/A2)	A2 Case AF	A2 Control AF	p-Value
2052146	160	71227260	A/C	0.010	0.042	0.0014
2052145	6053	71233153	T/G	0.858	0.776	0.0007
740980	9719	71236819	A/G	0.620	0.644	0.4134
758915	10481	71237581	T/C	0.718	0.718	0.9903
758914	10676	71237776	A/T	0.754	0.749	0.8560
2098195	17179	71244279	C/G	0.976	0.989	0.1034
740979	18561	71245661	A/T	0.656	0.694	0.1850

dbSNP rs#	Position in Fig 3	Chrom Position	Alleles (A1/A2)	A2 Case AF	A2 Control AF	p-Value
740978	18658	71245758	G/C	0.011	0.047	0.0005
740977	18694	71245794	A/G	0.913	0.873	0.0353
740976	18858	71245958	T/C	0.610	0.676	0.0217
2052143	24582	71251682	G/A	0.466	0.405	0.0418
2052142	24683	71251783	G/A	0.015	0.051	0.0011
2052141	24767	71251867	C/T	0.363	0.315	0.0950
758913	27402	71254502	A/G	0.931	0.871	0.0011
740975	28150	71255250	T/G	0.461	0.514	0.0763
747987	28494	71255594	T/C	0.715	0.813	0.0003
1126160	32003	71259103	A/C	0.349	0.409	0.0392
2332918	35588	71262688	C/T	0.041	0.070	0.0355
2332919	35619	71262719	T/C	0.300	0.271	0.2797
1990443	35856	71262956	G/A	0.324	0.268	0.0407
3937455	36254	71263354	G/C	0.445	0.455	0.7518
973963	37314	71264414	G/A	0.029	0.035	0.6030
1990441	40033	71267133	T/G	0.128	0.152	0.2380
1990440	40095	71267195	G/C	0.744	0.842	0.0002
2159715	42593	71269693	A/C	0.534	0.542	0.7822
2109795	42799	71269899	A/G	0.795	0.747	0.0582
2159714	43090	71270190	G/A	0.035	0.036	0.9187
1468662	46683	71273783	A/G	0.035	0.069	0.0118
2215591	49774	71276874	A/G	0.892	0.857	0.0776
2109794	51796	71278896	C/T	0.042	0.041	0.9714
2877821	52079	71279179	A/T	0.778	0.862	0.0005
2191822	53857	71280957	T/C	0.899	0.845	0.0078
2191821	53971	71281071	A/C	0.427	0.422	0.8733
1544579	55899	71282999	T/C	0.496	0.483	0.6724
2215590	60682	71287782	G/A	0.271	0.285	0.5936
1004552	61291	71288391	C/T	0.393	0.378	0.5996
1860749	72720	71299820	G/A	0.652	0.522	0.0001
1860748	72752	71299852	A/C	0.894	0.820	0.0007
763388	85507	71312607	A/G	0.291	0.310	0.4883
1035099	89751	71316851	T/A	0.555	0.543	0.7079

[0263] Figure 15 shows the proximal SNPs in and around the DPF3 gene. As indicated, some of the SNPs were untyped. The position of each SNP on the chromosome is presented on the x-axis. The y-axis gives the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 15 can be determined by consulting Table 18. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0264] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). The black line (or generally top-most curve, *e.g.*, see peak in left-most

graph just to the left of position 92150000) provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0265] Finally, the gene or genes present in the loci region of the proximal SNPs as annotated by Locus Link (http address: www.ncbi.nlm.nih.gov/LocusLink/) are provided on the graph. The exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is place at the 3' end of each gene to show the direction of transcription.

Example 6

CENCP1 Proximal SNPs

[0266] It has been discovered that a polymorphic variation (rs355510) in the CENPC1 region is associated with the occurrence of breast cancer (see Examples 1 and 2). Subsequently, SNPs proximal to the incident SNP (rs355510) were identified and allelotyped in breast cancer sample sets and control sample sets as described in Examples 1 and 2. Approximately seventy-nine allelic variants located within the CENPC1 region were identified and allelotyped. The polymorphic variants are set forth in Table 19. The chromosome position provided in column four of Table 19 is based on Genome "Build 33" of NCBI's GenBank.

Table 19

dbSNP rs#	Chromosome	Position in Figure 4	Chromosome Position	Allele Variants
1874633	4	196	68275196	A/G
1846060	4	13311	68288311	G/A
451352	4	14486	68289486	C/T
355468	4	14691	68289691	A/T
355469	4	15551	68290551	C/G
355470	4	17702	68292702	T/C
355471	4	17872	68292872	T/C
191650	4	19588	68294588	T/C
355472	4	19910	68294910	T/A
1874635	4	20006	68295006	A/C
1497430	4	20575	68295575	A/G
2254659	4	21092	68296092	G/A
3822197	4	22830	68297830	C/T
2632453	4	23455	68298455	A/G
2646282	4	23716	68298716	G/A
2646285	4	23890	68298890	T/G
768244	4	24001	68299001	C/T
724199	4	24995	68299995	G/A

dbSNP rs#	Chromosome	Position in Figure 4	Chromosome Position	Allele Variants
1187960	4	27282	68302282	T/C
1187961	4	27779	68302779	C/T
355518	4	29099	68304099	C/G
355519	4	31185	68306185	A/G
355511	4	33994	68308994	C/T
451397	4	34942	68309942	T/C
355513	4	35137	68310137	C/G
355514	4	36538	68311538	T/C
355515	4	37139	68312139	C/T
1056789	4	37358	68312358	G/A
2646290	4	38828	68313828	A/G
190255	4	39469	68314469	T/C
355466	4	40233	68315233	T/C
355465	4	40472	68315472	A/T
2646292	4	41679	68316679	C/T
2632454	4	41682	68316682	G/A
1056787	4	42831	68317831	A/G
CENPC1_SNP1	4	42976	68317976	A/G
173317	4	44128	68319128	A/G
451344	4	44195	68319195	C/T
355510	4	46769	68321769	G/A
355508	4	47363	68322363	G/C
451391	4	48843	68323843	C/T
355500	4	52574	68327574	A/G
355499	4	52602	68327602	A/G
355498	4	53212	68328212	A/G
1187974	4	53781	68328781	C/G
355493	4	54710	68329710	A/T
2632456	4	55808	68330808	G/A
1825790	4	57987	68332987	T/A
355475	4	58556	68333556	C/A
1391110	4	59148	68334148	T/A
1442557	4	59286	68334286	G/C
355478	4	60217	68335217	A/G
189579	4	60412	68335412	G/T
355480	4	60753	68335753	C/T
355481	4	60791	68335791	T/G
355483	4	61524	68336524	A/G
355485	4	62543	68337543	T/C
2646267	4	62825	68337825	A/G
2646268	4	62826	68337826	A/C
355486	4	62857	68337857	C/T
355487	4	63400	68338400	T/C
355488	4	63960	68338960	T/A
355489	4	64307	68339307	A/G
451376	4	64539	68339539	A/G
1353626	4	65728	68340728	A/G
2632450	4	66000	68341000	G/A
2646269	4	66521	68341521	T/G
2276945	4	68185	68343185	C/T

dbSNP rs#	Chromosome	Position in Figure 4	Chromosome Position	Allele Variants
3775861	4	69643	68344643	G/A
1403151	4	74909	68349909	C/A
1843833	4	82973	68357973	T/G
1843831	4	83039	68358039	T/C
3806810	4	85713	68360713	A/G
3775862	4	86873	68361873	T/C
1962700	4	90293	68365293	T/G
2046601	4	91810	68366810	T/G
2171386	4	92609	68367609	A/G
2046599	4	92884	68367884	G/A
355490	4			A/T

Assay for Verifying and Allelotyping SNPs

[0267] The methods used to verify and allelotype the proximal SNPs of Table 19 are the same methods described in Examples 1 and 2 herein. The PCR primers and extend primers used in these assays are provided in Table 20 and Table 21, respectively.

Table 20

dbSNP rs#	Forward PCR primer	Reverse PCR primer
1056787	ACGTTGGATGCATTTCATATTTTGTAGATC	ACGTTGGATGTCTCAGCCCTCTGATAAAAC
1056789	ACGTTGGATGTGAAGGTTCTGGAGGTATCG	ACGTTGGATGTCTTCTTAGCCAAGTCTGCC
CENPC1_SNP1	ACGTTGGATGAACAACGCACAATATCCCCG	ACGTTGGATGGGGTGAGGTTTATGGGAATG
11250	ACGTTGGATGAACAACGCACAATATCCCCG	ACGTTGGATGCATTTGCCAAAGTCTTAGGT
1187960	ACGTTGGATGTGAACCCTTCAAAATCACCC	ACGTTGGATGTTGTGTTTCATGGGAGGAGG
1187961	ACGTTGGATGCAACAGATTTTCCCTGTAGAC	ACGTTGGATGTGCATTGACTTCTCCTCAGC
1187974	ACGTTGGATGGCTGAGCAGAAGCTCTTTCA	ACGTTGGATGTGGGCAAAGACTTCATGATT
1353626	ACGTTGGATGCAACTACTACCTAGATGATGA	ACGTTGGATGAATAGAAAATCTAAATTGTCTAC
1391110	ACGTTGGATGAGTATGAAGGTCAGGGTCAG	ACGTTGGATGAAAGAGCACTGACCATGGAG
1403151	ACGTTGGATGTCAGTCAGAGATCATAGTTC	ACGTTGGATGCATGTAGTGCTTTAACAAATG
1442557	ACGTTGGATGCAACACATGCACCATTAGCG	ACGTTGGATGGAAGCCACAAACAGATCAGG
1497430	ACGTTGGATGTTGCTTGCTTGATGATTGGC	ACGTTGGATGTCTTCTGGACTTTAGCACTG
173317	ACGTTGGATGCTATAGGACTGTAAATTGTAG	ACGTTGGATGTTTTTACACACATGCTGTCA
1825790	ACGTTGGATGGGCCAACATGGTAAAACTCC	ACGTTGGATGCTGGGATAACAGGTACTTGC
1843831	ACGTTGGATGTCTCAGCTCATTTCCACCTC	ACGTTGGATGACCTGTAGTCCCAGCTACTC
1843833	ACGTTGGATGGACCAACATGGTGAAATCTC	ACGTTGGATGTGAGTAGCTGGGACTACAGG
1846060	ACGTTGGATGAAGATTATCACCGCACTGGG	ACGTTGGATGATCTCCTGACCTCGTGATCC
1874633	ACGTTGGATGAGGTTTTTGGTATGGTTAGC	ACGTTGGATGGAAAAGGGAGTTGGCCTAAA
1874635	ACGTTGGATGAGAGAGAGAGAGAGAGAGAG	ACGTTGGATGATGGGCTATAGTGGGATAGG
189579	ACGTTGGATGACACCAAAAGCAATGGCAAC	ACGTTGGATGGTTGCCTGTTCACTCTGATG
190255	ACGTTGGATGGAGATCTAGCACATTTATCC	ACGTTGGATGAGGTTGCCTGAAATGCTAAG
191650	ACGTTGGATGGAGATACCTTTGCTAAGGTG	ACGTTGGATGGGTAGTAATAATGGTACTCC
1962700	ACGTTGGATGATAAGAGAGAGTGTGGGTGG	ACGTTGGATGATTTCCTGACCTCGTGATCC
2046599	ACGTTGGATGTATTGAATTCCCTCTGTATG	ACGTTGGATGTCATTCTTTTGAGACTGAAC
2046601	ACGTTGGATGGCTCCAATGACTAAGTGGAC	ACGTTGGATGGACAGAACACTAAGAGCCTA
2171386	ACGTTGGATGCTTATCGAAATGAAATCAAG	ACGTTGGATGACAGCTGCAAACCTAAGGAC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
2254659	ACGTTGGATGATCTCTAAGTGAGATAGAGG	ACGTTGGATGCCCAGTCAAATGAAACCCAC
2276945	ACGTTGGATGGGGAATTCTATATTCCCATTG	ACGTTGGATGCCCAATTCCAACAGAAAATATC
2632450	ACGTTGGATGTTGAGACAAGCCTAGGCAAC	ACGTTGGATGGTGCTGGGATTACAGGTGTG
2632453	ACGTTGGATGAAAAGTGAGAGGGCAATAGG	ACGTTGGATGCATAGTAAGTCACCACAAGC
2632454	ACGTTGGATGTTCTGTGGGTCAGATGTCTC	ACGTTGGATGAGAAACAGACTTCCTCCCAG
2632456	ACGTTGGATGCCACCATATCAACAGATCAG	ACGTTGGATGCCTGCCAGTATGCTGAGAAT
2646267	ACGTTGGATGTGAGAAAAAGCACTCCTGGG	ACGTTGGATGAGGCTGAGACAGGAGAATTG
2646268	ACGTTGGATGCAGGAGAATTGCTTGAACCC	ACGTTGGATGTGAGAAAAAGCACTCCTGGG
2646269	ACGTTGGATGACCACTATTGTTTCTTTCTC	ACGTTGGATGGGCTAAAGAGTGAAACCCTG
2646282	ACGTTGGATGGATTGTTTTGAGTCATCTAC	ACGTTGGATGCTGAAATTGACCAGGAAACAC
2646285	ACGTTGGATGGGTGGATTGGACAAACTTGC	ACGTTGGATGCCTTTTGCTTTTCATTGCTC
2646290	ACGTTGGATGGATAGCAAGCTACCTAAGAC	ACGTTGGATGCCTCCTTACTCCACTCAATC
2646292	ACGTTGGATGTTCTGTGGGTCAGATGTCTC	ACGTTGGATGCAAAGAAACAGACTTCCTCC
355465	ACGTTGGATGTATGAGGTTCTGCCACCAAG	ACGTTGGATGTACCAAATCTGAGGGTAGTC
355466	ACGTTGGATGCAGGAGCTGCTTAATTCCTC	ACGTTGGATGGATCTTGGGCACTAAGTCTC
355468	ACGTTGGATGCCTCTCCTCATTTCTGTAAAC	ACGTTGGATGGGCAGGTGGTTAGCATTAAG
355469	ACGTTGGATGTTGGGATCTAGGCATCAAGG	ACGTTGGATGAGGAGGCACATAATGCTTGG
355470	ACGTTGGATGACATACACACACACACACAC	ACGTTGGATGGAGACATACACCTCTGCAAC
355471	ACGTTGGATGCTCATTACAACTTCAGCCAG	ACGTTGGATGACTCAGGACTAAGCTAGTTG
355472	ACGTTGGATGTCTCTCTCTCTCTCTCTCTC	ACGTTGGATGCAGCCCTTAGTACTCAATGG
355475	ACGTTGGATGCTGTCTTATCCCAACTTAGA	ACGTTGGATGGTCATGTTACATACCGAAAC
355478	ACGTTGGATGGGAGGAATCCATATATAGGC	ACGTTGGATGCTGCTGAAGGGAATGAGTAC
355480	ACGTTGGATGGTTTACAGTCCCACCAACAG	ACGTTGGATGAGTCAGGAAACAACAGGTGC
355481	ACGTTGGATGATTGCCACACTGTCTTCCAC	ACGTTGGATGGGATGTGGAGAAACAGGAAC
355483	ACGTTGGATGCCATGTAAGTCTGTCATTTA	ACGTTGGATGAAGTGGTAGCAGAAGTGTGG
355485	ACGTTGGATGAAGAAGAGGCATGCAAACAG	ACGTTGGATGCTGCGACAAAAGACACATTC
355486	ACGTTGGATGTGAGAAAAAGCACTCCTGGG	ACGTTGGATGAGGAGAATTGCTTGAACCCG
355487	ACGTTGGATGCGAGGTAATGAGCAAAGTAAG	ACGTTGGATGGACATTAGGTTCATCTAACCC
355488	ACGTTGGATGCCAGTTTTCTATGACAAACG	ACGTTGGATGAAAGAGCAGGGACAGCAAAG
355489	ACGTTGGATGACTCTAGGTATTTTGACTCC	ACGTTGGATGAACTTCCATAGTAGAAAGCC
355490	ACGTTGGATGAACTTCCATAGTAGAAAGCC	ACGTTGGATGACTCTAGGTATTTTGACTCC
355493	ACGTTGGATGAGTGGTTTGCTGCACCTATC	ACGTTGGATGGGGAGAGCATTAGGACAAAC
355498	ACGTTGGATGATGAGAGAGGACACAAAGAG	ACGTTGGATGTTACTTTGCACAGTGTGGCC
355499	ACGTTGGATGCAATCAAGCAGAAGGATGGG	ACGTTGGATGGGTGTCTTCTTATAGTTGTC
355500	ACGTTGGATGCAATCAAGCAGAAGGATGGG	ACGTTGGATGGGTGTCTTCTTATAGTTGTC
355508	ACGTTGGATGGTGTAGATGTGTATCAGGTCA	ACGTTGGATGGTCCACAAAGCATAGCATCC
355510	ACGTTGGATGCCCTCCTTTTAACCTTTTAGG	ACGTTGGATGTTCTGAGATGATCCTGATGG
355511	ACGTTGGATGCAGGAGGATATGTGAAAGTC	ACGTTGGATGGTGGATACCAAAATCCAAGG
355513	ACGTTGGATGTGCTGTATAACAGATTACCC	ACGTTGGATGAACTAGCTAGCTAAGCCTCC
355514	ACGTTGGATGCCTCAATAGGTTGTTGGAAC	ACGTTGGATGTTGAGTTCATACTATGTGCC
355515	ACGTTGGATGAGCTCTGCACTCTGACATAC	ACGTTGGATGGTGCAGAGTACTACTTTGCC
355518	ACGTTGGATGTGCCATGGGGTTGTAAAATC	ACGTTGGATGACACAGAGACCAGCTGAAAG
355519	ACGTTGGATGGGGAAGAAGCAGATTTTGAG	ACGTTGGATGCATAGGTTGAGAACATCAAGC
3775861	ACGTTGGATGCCATCTCTTTGAAAATTCCAC	ACGTTGGATGCCCTCAAGTACTTGTTTTGTC
3775862	ACGTTGGATGTAATGAAGCTGAGTTTATTC	ACGTTGGATGGTTTTTGTTTATTGGTGTCC
3806810	ACGTTGGATGTCTTTTCTCCCATCATTTCC	ACGTTGGATGACTCAATGGTTGCATGTAGG
3822197	ACGTTGGATGTGTTTGCTAAAGCTATGCTG	ACGTTGGATGTGAGCATTATGCCTAAGAGC
451344	ACGTTGGATGCCTTTCTAGATACACTCCAT	ACGTTGGATGCAGCATGTGTGTAAAAATGC
451352	ACGTTGGATGAGGCAAATTATTTTTGGATG	ACGTTGGATGCTCCCTAAATGGGGAAAAAAG
451362	ACGTTGGATGCAACACATGCACCATTAGCG	ACGTTGGATGGAAGCCACAAACAGATCAGG

dbSNP rs#	Forward PCR primer	Reverse PCR primer
451376	ACGTTGGATGAGCAGTCTATTCTGGTTCAC	ACGTTGGATGGCCTTTGAGCTTTAAAAATC
451391	ACGTTGGATGTAAAGTAGGGACTGGGATGG	ACGTTGGATGGCTGTAGAGTAGTGAAACCC
451397	ACGTTGGATGGTTGCCATATTCAGCAGCTG	ACGTTGGATGCTGTTTCCAGTAGACCTTAG
724199	ACGTTGGATGCCAGCTAAAACTGCAAATAC	ACGTTGGATGTGGACTCATTTGAGAATATG
768244	ACGTTGGATGTAAAACCCCTTCCTCATCCC	ACGTTGGATGACCTTTAGCAGCCTGAAACC

Table 21

dbSNP rs#	Extend Primer	Term Mix
355469	GCACATAATGCTTGGTTGTATT	ACT
CENPC1_SNP1	CTTGACTTTCTACCTTGAA	ACT
11250	CTCTTGACTTTCTACCTTGAA	ACT
173317	ACTTAGCGGCTTAAAACAAC	ACT
189579	CTGTTCACTCTGATGGTAGTTT	CGT
190255	GTACTATGTGGCAGATGA	ACT
191650	GGTACTCCTACTTAAAATTTTG	ACT
355465	GAGGGTAGTCTTGGGAACC	CGT
355466	CTCTAGTGAGCTTCCCT	ACT
355468	AGCATTAAGTATTCATGAGAGTTC	CGT
355470	GGTCTGTTTTATATGTGTGT	ACT
355471	AGCTAGTTGCTTCAGTAAGT	ACT
355472	GTACAGTCATAACAGTTGTTAA	CGT
355475	TACATACCGAAACACATTCC	CGT
355478	ACATTCTATATGGCCCCTTG	ACT
355480	GGAGAGGATGTGGAGAAA	ACG
355481	GGTGGGACTGTAAACTA	ACT
355483	AGAAGTGTGGACACAGTATC	ACT
355485	CACATTCAACTATACACGCTTTTA	ACT
355486	GTGAGCCGAAATCGTGCCAC	ACG
355487	TTCATCTAACCCTTTTCATAA	ACT
355488	AGCAAAGCTGAAAATGATAA	CGT
355489	CAATAAATAATAGCAAAGACTGG	ACT
355490	TGTTTATATTGCTGTTTCTTGA	CGT
355493	CTCATGTGGGGCTTAAA	CGT
355498	GTGTGGCCATTTTCACT	ACT
355499	TGTTAGATAGAGGTTTATCATTTT	ACT
355500	TTTTTCCTGCAATAGTTTTCT	ACT
355508	ATACTTATGCTCTGCTACC	ACT
355510	ATGGTTTTCTTTCTTGTCCTTC	ACG
355511	GGATGCTCAAGTCCCTTATATA	ACG
355513	GCCTCCCAGATTGCTGA	ACT
355514	TGTGCCAAATATTTGCTAGAT	ACT
355515	ACTACTTTGCCTGTGTGTCA	ACG
355518	ACCAGCTGAAAGAAAATC	ACT
355519	AAGCTTAGTATGTCCAAATCTAAC	ACT
451344	GTGTGTAAAAATGCATTCCAAGTT	ACG

dbSNP rs#	Extend Primer	Term Mix
451352	CCCCCGAAATGTTTCAAAGG	ACG
451362	CCACAAACAGATCAGGTTGGTG	ACT
451376	AGTATGTAAAAAGATAGGGAAGA	ACT
451391	GAGTAGTGAAACCCCTGACC	ACG
451397	CAGTAGACCTTAGTTTCTTAACC	ACT
724199	GAGAATATGATAAAAGCTCAGACC	ACG
768244	GTTTCTGTCTCTGGCGA	ACG
1056787	GGATACAAGTTATGCTTTGATAG	ACT
1056789	TCCAATGGCTCACTCAG	ACG
1187960	GGAGGAGGTCAAAATATCA	ACT
1187961	GACTTCTCCTCAGCTATGAA	ACG
1187974	TGATTAAAACACCAAAAGCAATT	ACT
1353626	AATCTAAATTGTCTACTGAAACT	ACT
1391110	CCATGGAGTTGTAAGGAA	CGT
1403151	TAGTGCTTTAACAAATGCTGTCA	CGT
1442557	CACAAACAGATCAGGTTGGTG	ACT
1497430	GAATTGGGGAGAGAAAGGGA	ACT
1825790	CCTGGCAAATTTTGGTATTTTTAG	CGT
1843831	GCGGGAGAATGGCATGA	ACT
1843833	GCTCACCACCACACCTG	ACT
1846060	AAAGTGCTGGGATTACAGG	ACG
1874633	TGGCCTAAAAATATTTTTACCGT	ACT
1874635	CAACTGTTTAACAACCAGGC	ACT
1962700	AGAGTGCTGGGATTACA	ACT
2046599	CTTTTGAGACTGAACACCTCTA	ACG
2046601	AGAACACTAAGAGCCTAGAATGG	ACT
2171386	AGTATGCAGAGACTTACAG	ACT
2254659	AACCCACCATTCCTATG	ACG
2276945	CACAAAATACCTCCAAATTTTA	ACG
2632450	TTACAGGTGTGAGCCAC	ACG
2632453	CACCACAAGCCACTTGA	ACT
2632454	CTTCCTCCCAGAGCCAC	ACG
2632456	TCATAGGTAATGTGGATTTTGT	ACG
2646267	TTGCTTGAACCCGGGAG	ACT
2646268	TCGGCTCACTGCAATCTCT	ACT
2646269	TTCTCGCAAAGAGAAAAC	ACT
2646282	GGAATTAGCAGTCATTTCTTA	ACG
2646285	ATTTCTCTAGACTTTGCTACAAT	ACT
2646290	AGTTCATCCTTCAGGAA	ACT
2646292	AGACTTCCTCCCAGAGC	ACG
3775861	GTTTTGTCTTCAAATAGTAAAGA	ACG
3775862	TCCATTTTTATTTGCAGAAGAC	ACT
3806810	ATTGGATTTGGCGTAGC	ACT
3822197	AGCAGTAGGCAACTTCT	ACG

Genetic Analysis of Allelotyping Results

[0268] Allelotyping results are shown for cases and controls in Table 22. The allele frequency for the A2 allele is noted in the fifth and sixth columns for breast cancer pools and control pools, respectively, where "AF" is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs1874633 has the following case and control allele frequencies: case A1 (A) = 0.514; case A2 (G) = 0.486; control A1 (A) = 0.449; and control A2 (G) = 0.551, where the nucleotide is provided in paranthesis. SNPs with blank allele frequencies were untyped.

Table 22

dbSNP rs#	Position in Figure 4	Chromosome Position	A1/A2 Allele	A2 Case AF	A2 Control AF	p-Value
1874633	196	68275196	A/G	0.486	0.551	0.0292
1846060	13311	68288311	G/A	0.416	0.468	0.0792
451352	14486	68289486	C/T	0.474	0.411	0.0365
355468	14691	68289691	A/T	0.839	0.839	0.9913
355469	15551	68290551	C/G	0.089	0.072	0.3028
355470	17702	68292702	T/C	0.077	0.059	0.2261
355471	17872	68292872	T/C	0.476	0.442	0.2613
191650	19588	68294588	T/C	0.122	0.103	0.3282
355472	19910	68294910	T/A	0.491	0.568	0.0114
1874635	20006	68295006	A/C	0.206	0.238	0.2083
1497430	20575	68295575	A/G	0.389	0.476	0.0039
2254659	21092	68296092	G/A	0.554	0.587	0.2664
3822197	22830	68297830	C/T	0.028	0.018	0.2999
2632453	23455	68298455	A/G	0.866	0.895	0.1407
2646282	23716	68298716	G/A	0.137	0.090	0.0146
2646285	23890	68298890	T/G	0.400	0.335	0.0269
768244	24001	68299001	C/T	0.299	0.286	0.6333
724199	24995	68299995	G/A	0.446	0.374	0.0150
1187960	27282	68302282	T/C	0.071	0.060	0.4859
1187961	27779	68302779	C/T	0.499	0.549	0.0968
355518	29099	68304099	C/G	0.432	0.491	0.0473
355519	31185	68306185	A/G	0.095	0.076	0.2836
355511	33994	68308994	C/T	0.450	0.361	0.0030
451397	34942	68309942	T/C	0.442	0.512	0.0210
355513	35137	68310137	C/G	0.385	0.334	0.0748
355514	36538	68311538	T/C	0.423	0.479	0.0596
355515	37139	68312139	C/T	0.422	0.362	0.0395
1056789	37358	68312358	G/A	0.494	0.539	0.1409
2646290	38828	68313828	A/G	0.393	0.337	0.0559
190255	39469	68314469	T/C	0.459	0.514	0.0664
355466	40233	68315233	T/C	0.404	0.468	0.0328
355465	40472	68315472	A/T	0.481	0.547	0.0281
2646292	41679	68316679	C/T	0.422	0.370	0.0820
2632454	41682	68316682	G/A	0.914	0.936	0.1705
1056787	42831	68317831	A/G	0.909	0.860	0.0112
CENPC1_SNP1	42976	68317976	A/G	0.367	0.306	0.0322
173317	44128	68319128	A/G	0.087	0.080	0.6745
451344	44195	68319195	C/T	0.366	0.307	0.0392
355510	46769	68321769	G/A	0.487	0.514	0.3645
355508	47363	68322363	G/C	0.086	0.070	0.3357
451391	48843	68323843	C/T	0.440	0.370	0.0171
355500	52574	68327574	A/G	0.874	0.904	0.1103
355499	52602	68327602	A/G	0.874	0.884	0.5959
355498	53212	68328212	A/G	0.477	0.528	0.0932

dbSNP rs#	Position in Figure 4	Chromosome Position	A1/A2 Allele	A2 Case AF	A2 Control AF	p-Value
1187974	53781	68328781	C/G	0.563	0.540	0.4558
355493	54710	68329710	A/T	0.950	0.932	0.2013
2632456	55808	68330808	G/A	0.091	0.074	0.3234
1825790	57987	68332987	T/A	0.043	0.067	0.0709
355475	58556	68333556	C/A	0.252	0.199	0.0343
1391110	59148	68334148	T/A	0.696	0.679	0.5418
1442557	59286	68334286	G/C	0.458	0.523	0.0306
355478	60217	68335217	A/G	0.314	0.371	0.0474
189579	60412	68335412	G/T	0.008	0.002	0.1543
355480	60753	68335753	C/T	0.905	0.910	0.7624
355481	60791	68335791	T/G	0.974	0.979	0.5823
355483	61524	68336524	A/G	0.371	0.414	0.1461
355485	62543	68337543	T/C	0.487	0.541	0.0732
2646267	62825	68337825	A/G	0.368	0.312	0.0520
2646268	62826	68337826	A/C	0.306	0.239	0.0123
355486	62857	68337857	C/T	0.438	0.375	0.0316
355487	63400	68338400	T/C	0.468	0.559	0.0031
355488	63960	68338960	T/A	0.533	0.454	0.0090
355489	64307	68339307	A/G	0.367	0.324	0.1291
451376	64539	68339539	A/G	0.873	0.871	0.9287
1353626	65728	68340728	A/G	0.356	0.383	0.3657
2632450	66000	68341000	G/A	0.256	0.259	0.9210
2646269	66521	68341521	T/G	0.084	0.062	0.1648
2276945	68185	68343185	C/T	0.459	0.510	0.0866
3775861	69643	68344643	G/A	0.532	0.521	0.7150
1403151	74909	68349909	C/A	0.739	0.801	0.0148
1843833	82973	68357973	T/G	0.920	0.939	0.2355
1843831	83039	68358039	T/C	0.032	0.040	0.5196
3806810	85713	68360713	A/G	0.078	0.058	0.1942
3775862	86873	68361873	T/C	0.744	0.765	0.4224
1962700	90293	68365293	T/G	0.733	0.739	0.8308
2046601	91810	68366810	T/G	0.080	0.073	0.6571
2171386	92609	68367609	A/G	0.685	0.662	0.4056
2046599	92884	68367884	G/A	0.717	0.755	0.1540
355490			A/T	0.495	0.548	0.0763

[0269] Figure 16 shows the proximal SNPs in and around the *ICAM* region for females. The position of each SNP on the chromosome is presented on the x-axis. The y-axis gives the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 16 can be determined by consulting Table 22. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0270] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie,

Wadsworth & Brooks/Cole.). The black line (or generally top-most curve, *e.g.*, see peak in left-most graph just to the left of position 92150000) provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0271] Finally, the gene or genes present in the loci region of the proximal SNPs as annotated by Locus Link (http address: www.ncbi.nlm.nih.gov/LocusLink/) are provided on the graph. The exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is place at the 3' end of each gene to show the direction of transcription.

Additional Genotyping

[0272] In addition to the CENCP1 incident SNP, another SNP (rs1056787) was genotyped in the discovery cohort and found to be significantly associated with breast cancer with a p-value of 0.0266. See Table 25.

[0273] The methods used to verify and genotype the proximal SNP of Table 15 are the same methods described in Examples 1 and 2 herein. The PCR primers and extend primers used in these assays are provided in Table 11 and Table 12, respectively.

Table 23

dbSNP rs#	Second PCR primer	First PCR primer
1056787	ACGTTGGATGCATTTCATATTTTGTAGATC	ACGTTGGATGTCTCAGCCCTCTGATAAAAC

Table 24

dbSNP rs#	Extend Primer	Term Mix
1056787	GGATACAAGTTATGCTTTGATAG	ACT

[0274] Table 13, below, shows the case and control allele frequencies along with the p-values for the SNPs genotyped. The disease associated allele of column 4 is in bold and the disease associated amino acid of column 5 is also in bold. The chromosome position provided corresponds to NCBI's Build 33.

Table 25: Genotpying Results

dbSNP rs#	Position in Figure 4	Chromo-some Position	Alleles (A1/A2)	Amino Acid Change	AF F case	AF F control	p-value	Odds Ratio
1056787	42831	68317831	A/G	D389G	A = 0.030 G = 0.970	A = 0.110 G = 0.890	**0.0266**	1.640

Example 7

In Vitro Production of Target Polypeptides

[0275] cDNA is cloned into a pIVEX 2.3-MCS vector (Roche Biochem) using a directional cloning method. A cDNA insert is prepared using PCR with forward and reverse primers having 5' restriction site tags (in frame) and 5-6 additional nucleotides in addition to 3' gene-specific portions, the latter of which is typically about twenty to about twenty-five base pairs in length. A Sal I restriction site is introduced by the forward primer and a Sma I restriction site is introduced by the reverse primer. The ends of PCR products are cut with the corresponding restriction enzymes (*i.e.*, Sal I and Sma I) and the products are gel-purified. The pIVEX 2.3-MCS vector is linearized using the same restriction enzymes, and the fragment with the correct sized fragment is isolated by gel-purification. Purified PCR product is ligated into the linearized pIVEX 2.3-MCS vector and *E. coli* cells transformed for plasmid amplification. The newly constructed expression vector is verified by restriction mapping and used for protein production.

[0276] *E. coli* lysate is reconstituted with 0.25 ml of Reconstitution Buffer, the Reaction Mix is reconstituted with 0.8 ml of Reconstitution Buffer; the Feeding Mix is reconstituted with 10.5 ml of Reconstitution Buffer; and the Energy Mix is reconstituted with 0.6 ml of Reconstitution Buffer. 0.5 ml of the Energy Mix was added to the Feeding Mix to obtain the Feeding Solution. 0.75 ml of Reaction Mix, 50 µl of Energy Mix, and 10 µg of the template DNA is added to the *E. coli* lysate.

[0277] Using the reaction device (Roche Biochem), 1 ml of the Reaction Solution is loaded into the reaction compartment. The reaction device is turned upside-down and 10 ml of the Feeding Solution is loaded into the feeding compartment. All lids are closed and the reaction device is loaded into the RTS500 instrument. The instrument is run at 30°C for 24 hours with a stir bar speed of 150 rpm. The pIVEX 2.3 MCS vector includes a nucleotide sequence that encodes six consecutive histidine amino acids on the C-terminal end of the target polypeptide for the purpose of protein purification. Target polypeptide is purified by contacting the contents of reaction device with resin modified with Ni^{2+} ions. Target polypeptide is eluted from the resin with a solution containing free Ni^{2+} ions.

Example 8

Cellular Production of Target Polypeptides

[0278] Nucleic acids are cloned into DNA plasmids having phage recombination cites and target polypeptides are expressed therefrom in a variety of host cells. Alpha phage genomic DNA contains short sequences known as attP sites, and *E. coli* genomic DNA contains unique, short sequences known as attB sites. These regions share homology, allowing for integration of phage DNA into *E. coli* via directional, site-specific recombination using the phage protein Int and the *E. coli* protein IHF. Integration produces two new att sites, L and R, which flank the inserted prophage DNA. Phage excision from *E. coli* genomic DNA can also be accomplished using these two proteins with the addition of a second phage protein, Xis. DNA vectors have been produced where the integration/excision process is modified to allow for the directional integration or excision of a target DNA fragment into a backbone vector in a rapid *in vitro* reaction (Gateway™ Technology (Invitrogen, Inc.)).

[0279] A first step is to transfer the nucleic acid insert into a shuttle vector that contains attL sites surrounding the negative selection gene, ccdB (*e.g.* pENTER vector, Invitrogen, Inc.). This transfer process is accomplished by digesting the nucleic acid from a DNA vector used for sequencing, and to ligate it into the multicloning site of the shuttle vector, which will place it between the two attL sites while removing the negative selection gene ccdB. A second method is to amplify the nucleic acid by the polymerase chain reaction (PCR) with primers containing attB sites. The amplified fragment then is integrated into the shuttle vector using Int and IHF. A third method is to utilize a topoisomerase-mediated process, in which the nucleic acid is amplified via PCR using gene-specific primers with the 5' upstream primer containing an additional CACC sequence (*e.g.*, TOPO® expression kit (Invitrogen, Inc.)). In conjunction with Topoisomerase I, the PCR amplified fragment can be cloned into the shuttle vector via the attL sites in the correct orientation.

[0280] Once the nucleic acid is transferred into the shuttle vector, it can be cloned into an expression vector having attR sites. Several vectors containing attR sites for expression of target polypeptide as a native polypeptide, N-fusion polypeptide, and C-fusion polypeptides are commercially available (*e.g.*, pDEST (Invitrogen, Inc.)), and any vector can be converted into an expression vector for receiving a nucleic acid from the shuttle vector by introducing an insert having an attR site flanked by an antibiotic resistant gene for selection using the standard methods described above. Transfer of the nucleic acid from the shuttle vector is accomplished by directional recombination using Int, IHF, and Xis (LR clonase). Then the desired sequence can be transferred to an expression vector by carrying out a one hour incubation at room temperature with Int, IHF, and Xis, a ten minute incubation at 37°C with proteinase K, transforming bacteria and allowing expression for one hour, and then plating on selective media. Generally, 90% cloning efficiency is achieved by this method. Examples of expression vectors are pDEST 14 bacterial expression vector with att7

promoter, pDEST 15 bacterial expression vector with a T7 promoter and a N-terminal GST tag, pDEST 17 bacterial vector with a T7 promoter and a N-terminal polyhistidine affinity tag, and pDEST 12.2 mammalian expression vector with a CMV promoter and neo resistance gene. These expression vectors or others like them are transformed or transfected into cells for expression of the target polypeptide or polypeptide variants. These expression vectors are often transfected, for example, into murine-transformed a adipocyte cell line 3T3-L1, (ATCC), human embryonic kidney cell line 293, and rat cardiomyocyte cell line H9C2.

Example 9

Haplotype analysis of the *KIAA0783* locus

[0281] Markers rs1681290, rs220097, rs3801435, and rs2883140 are significantly associated with breast cancer at the allele and genotype levels ($P < 0.05$). Strong LD is observed between markers 1681290, 220097, 3801435, and 2883140 ($r^2 > 0.90$). Pearson chi-squared statistics indicate that haplotypes are significantly associated with breast cancer. Haplotypes TTGCGG, CTGCGG, and TCATAT contribute most to the aggregate test statistic. Odds ratios and score tests indicate that individuals with the TTGCGG and CTGCGG haplotypes are significantly less likely to have breast cancer, while individuals with the TCATAT haplotype are slightly more likely to be affected than individuals with other haplotypes.

Statistics

[0282] Chi-squared statistics are estimated to assess whether 1) alleles and genotypes are associated with breast cancer status and 2) marker genotype frequencies deviate significantly from Hardy-Weinberg equilibrium (HWE). Haplotype frequencies and relative frequencies are estimated, as well as several statistics (r^2, D', and p-value) that gauge the extent and stability of linkage disequilibrium between markers in each region. Chi-squared statistics and score tests are estimated to determine whether reconstructed haplotypes are significantly associated with breast cancer status ($P < 0.05$). P-values are estimated for 1) the full set of reconstructed haplotypes and 2) a reduced set that excludes haplotypes with observed frequencies less than 10. Results are presented by chromosome order.

Results .

Summary Statistics: Alleles and Genotypes

SNP Locations

SNP.ID	Type	Location
218981	Proximal	10720511
1681284	Proximal	10739011
1681290	Proximal	10741656
220097	Incident	10759860
3801435	Proximal	10771563
2883140	Proximal	10806368

Allele by GYNGroup

	N	Case (N=510)	Control (N=538)	Test Statistic
218981:T	1028	47%(232)	45%(239)	Chi-square=0.68 d.f.=1 P=0.41
1681284:C	1032	56%(276)	50%(267)	Chi-square=3.51 d.f.=1 P=0.0608
1681290:A	1018	72%(352)	63%(330)	Chi-square=8.92 d.f.=1=0.00282
220097:C	996	29%(139)	38%(196)	Chi-square=8.03 d.f.=1P=0.00461|
3801435:G	1018	28%(138)	38%(200)	Chi-square=9.69 d.f.=1P=0.00185|
2883140:T	1012	73%(351)	62%(330)	Chi-square=12.78 d.f.=1 P<0.001|

Genotype by GYNGroup

	N	Case (N=255)	Control (N=269)	Test Statistic
218981:CC	514	27%(67)	27%(73)	Chi-square=2.41 d.f.=2 P=0.299
CT		51%(126)	56%(151)	
TT		22%(53)	16%(44)	
1681284:TT	516	19%(48)	26%(70)	Chi-square=3.77 d.f.=2 P=0.152
TC		50%(124)	48%(129)	
CC		31%(76)	26%(69)	
1681290:GG	509	9%(21)	16%(41)	Chi-square=8.64 d.f.=2 P=0.0133

	N	Case (N=255)	Control (N=269)	Test Statistic
GA		40%(98)	43%(114)	
AA		52%(127)	41%(108)	
220097:TT	498	50%(119)	40%(104)	Chi-square=8.06 d.f.=2 P=0.0177
TC		42%(99)	45%(116)	
CC		8%(20)	15%(40)	
3801435:AA	509	51%(124)	40%(107)	Chi-square=9.78 d.f.=2 P=0.0075
AG		41%(100)	44%(118)	
GG		8%(19)	15%(41)	
2883140:GG	506	8%(19)	16%(42)	Chi-square=12.14 d.f.=2 P=0.00231
GT		39%(93)	44%(116)	
TT		54%(129)	40%(107)	

Genotype QC: Test of Hardy-Weinberg Proportions

All

	A.freq	D	ChiSq	Pvalue
218981	0.543	-0.01990	3.290	0.0697
1681284	0.526	0.00564	0.263	0.6080
1681290	0.670	0.01170	1.430	0.2320
220097	0.664	0.00584	.351	.5530
3801435	0.667	0.00585	.355	.5510
2883140	0.675	.01360	.970	.1610

Control

	A.freq	D	ChiSq	Pvalue
218981	0.554	-0.03380	5.010	0.0252
1681284	0.502	0.01030	0.453	0.5010
1681290	0.627	0.01470	1.050	0.3050
220097	0.620	0.00904	0.393	0.5310
3801435	0.624	0.01190	0.684	0.4080
2883140	0.625	0.01700	1.410	0.2350

Summary Statistics: Linkage Disequilibrium

PHASE Haplotype Frequencies

	H.freq	H.relfreq
CCATAT	91	0.089
CCGCGG	4	0.004
CTACGG	5	0.005
CTACGT	1	0.001
CTATAT	142	0.138
CTGCAG	1	0.001
CTGCAT	2	0.002
CTGCGG	300	0.292
CTGCGT	10	0.010
CTGTAT	1	0.001
TCACGG	1	0.001
TCATAG	1	0.001
TCATAT	443	0.432
TTATAT	3	0.003
TTGCGG	21	0.020

Linkage Disequilibrium Between Markers

r^2

x	218981	1681284	1681290	220097	3801435	2883140
218981	1.000	0.603	0.311	0.316	0.311	0.292
1681284	0.603	1.000	0.524	0.532	0.525	0.498
1681290	0.311	0.524	1.000	0.965	0.952	0.914
220097	0.316	0.532	0.965	1.000	0.987	0.940
3801435	0.311	0.525	0.952	0.987	1.000	0.944
2883140	0.292	0.498	0.914	0.940	0.944	1.000

D'

	218981	1681284	1681290	220097	3801435	2883140
218981	1.000	0.803	0.728	0.725	0.724	0.715
1681284	0.803	1.000	0.978	0.972	0.972	0.966
1681290	0.728	0.978	1.000	0.996	0.982	0.969
220097	0.725	0.972	0.996	1.000	1.000	0.995
3801435	0.724	0.972	0.982	1.000	1.000	0.991
2883140	0.715	0.966	0.969	0.995	0.991	1.000

P-value

	218981	1681284	1681290	220097	3801435	2883140
218981	1	0	0	0	0	0
1681284	0	1	0	0	0	0
1681290	0	0	1	0	0	0
220097	0	0	0	1	0	0
3801435	0	0	0	0	1	0
2883140	0	0	0	0	0	1

Haplotype by GYNGroup

All Haplotypes

	Case	Case(%)	Case.X^2	Control	Control(%)	Control.X^2	OR	ln.OR
CTGCAG	0	0.00	0.48	1	0.10	0.44	0.0000	-Inf
TCACGG	0	0.00	0.48	1	0.10	0.44	-Inf	
						0.0000		
TCATAG	0	0.00	0.48	1	0.10	0.44	0.0000	-Inf
TTATAT	0	0.00	1.44	3	0.29	1.33	0.0000	-Inf
TTGCGG	1	0.10	8.17	20	1.95	7.53	0.0491	-3.0139
CCGCGG	1	0.10	0.44	3	0.29	0.40	0.3327	-1.1005
CTACGG	2	0.19	0.07	3	0.29	0.06	0.6660	-0.4065
CTGCGG	129	12.57	1.53	171	16.67	1.41	0.7191	-0.3298
CCATAT	43	4.19	0.01	48	4.68	0.01	0.8913	-0.1151

	Case	Case(%)	Case.X^2	Control	Control(%)	Control.X^2	OR	ln.OR
CTGCAT	1	0.10	0.00	1	0.10	0.00	1.0000	0.0000
TCATAT	230	22.42	1.45	213	20.76	1.34	1.1029	0.0979
CTATAT	76	7.41	0.92	66	6.43	0.85	1.1636	0.1515
CTGCGT	7	0.68	1.01	3	0.29	0.93	2.3425	0.8512
CTACGT	1	0.10	0.56	0	0.00	0.52	Inf	Inf
CTGTAT	1	0.10	0.56	0	0.00	0.52	Inf	Inf

Pearson Chi-squared Test = 33.8392, DF = 14, P-value = 0.002177

Permutation Test P-value = 0.01

PHASE Haplotypes (Low Frequency Excluded)

	Case	Case(%)	Case.X^2	Control	Control(%)	Control.X^2	OR	ln.OR
TTGCGG	1	0.10	8.23	20	1.99	7.68	0.0491	-3.0139
CTGCGG	129	12.81	1.72	171	16.98	1.61	0.7183	-0.3309
CCATAT	43	4.27	0.02	48	4.77	0.02	0.8912	-0.1152
TCATAT	230	22.84	1.23	213	21.15	1.14	1.1034	0.0984
CTATAT	76	7.55	0.81	66	6.55	0.76	1.1639	0.1518
CTGCGT	7	0.70	0.98	3	0.30	0.91	2.3427	0.8513

Pearson Chi-squared Test = 25.1157, DF = 5, P-value = 0.0001323

haplo.score Haplotypes

	Hap.Freq	Score	P. X^2	P.Sim
TTGCGG	0.0203	-3.7664	0.0002	0.0001
TTATAT	0.0063	-2.5040	0.0123	0.0097
CTGCGG	0.2947	-2.0103	0.0444	0.0438
CCATAT	0.0902	-0.3982	0.6905	0.7174
CTATAT	0.1318	1.4254	0.1540	0.1538
CTGCGT	0.0084	1.5778	0.1146	0.1243
TCATAT	0.4342	2.3889	0.0169	0.0180

Global Score = 27.2432, DF = 7, Global P.X^2 = 3e-04, Global P.Sim = 1e-04

Example 10

Haplotype analysis of the *CENPC1* locus

[0283] Each SNP noted below is significantly associated with breast cancer at allele level ($P < 0.05$). rs355510 maintains a significant relationship with disease at the genotype level. Near-complete LD is observed across the entire region. Pearson chi-squared statistics demonstrate that haplotypes CCAC and TTGT are significantly associated with breast cancer after low frequency haplotypes are removed from the analysis. Odds ratios and score tests indicate that individuals with the CCAC haplotype are significantly less likely to have breast cancer, while individuals with the TTGT haplotype are at moderately increased risk for disease vs. individuals with other haplotypes.

Statistics

[0284] Chi-squared statistics are estimated to assess whether 1) alleles and genotypes are associated with breast cancer status and 2) marker genotype frequencies deviate significantly from Hardy-Weinberg equilibrium (HWE). Haplotype frequencies and relative frequencies are estimated, as well as several statistics (r^2, D', and p-value) that gauge the extent and stability of linkage disequilibrium between markers in each region. Chi-squared statistics and score tests are estimated to determine whether reconstructed haplotypes are significantly associated with breast cancer status ($P < 0.05$). P-values are estimated for 1) the full set of reconstructed haplotypes and 2) a reduced set that excludes haplotypes with observed frequencies less than 10. Results are presented by chromosome order.

Results

Summary Statistics: Alleles and Genotypes

SNP Locations

SNP.ID	Type	Location
GP04.071927035	Proximal	68289486
355511	Proximal	68308994
355510	Incident	68321769
355487	Proximal	68338400

Allele by GYNGroup

	N	Case (N=508)	Control (N=536)	Test Statistic

Genotype by GYNGroup

	N	Case (N=254)	Control (N=268)	Test Statistic
GP04.071927035:CC	511	28%(69)	37%(98)	Chi-square=5.33 d.f.=2 P=0.0695
CT		52%(127)	48%(129)	
TT		20%(49)	15%(39)	
355511:TT	505	20%(48)	14%(38)	Chi-square=4.47 d.f.=2 P=0.107
TC		51%(124)	49%(129)	
CC		29%(70)	37%(96)	
355510:GG	502	20%(49)	15%(38)	Chi-square=6.52 d.f.=2 P=0.0383
GA		52%(125)	47%(123)	
AA		28%(68)	38%(99)	
355487:TT	496	20%(48)	15%(37)	Chi-square=5.35 d.f.=2 P=0.069
TC		52%(126)	48%(123)	
CC		28%(68)	37%(94)	

Genotype QC: Test of Hardy-Weinberg Proportions

All

	A.freq	D	ChiSq	Pvalue
GP04.071927035	0.577	-0.00599	0.303	0.582
355511	0.579	-0.00630	0.337	0.562
355510	0.577	-0.00599	0.303	0.582
355487	0.577	-0.00599	0.303	0.582

Control

	A.freq	D	ChiSq	Pvalue
GP04.071927035	0.609	-0.00420	0.0814	0.775
355511	0.611-0.00653	0.1970	0.657	
355510	0.609	-0.00420	0.0814	0.775
355487	0.611	-0.00271	0.0340	0.854

Summary Statistics: Linkage Disequilibrium

PHASE Haplotype Frequencies

	H.freq	H.relfreq
CCAC	581	0.576
CCAT	1	0.001
TCGT	2	0.002
TTGC	1	0.001
TTGT	423	0.420

Linkage Disequilibrium Between Markers

r^2

	GP04.071927035	355511	355510	355487
GP04.071927035	1.000	0.992	1.000	0.992
355511	0.992	1.000	0.992	0.984
355510	1.000	0.992	1.000	0.992
355487	0.992	0.984	0.992	1.000

D'

	GP04.071927035	355511	355510	355487
GP04.071927035	1.000	1.000	1.000	0.996
355511	1.000	1.000	1.000	0.996
355510	1.000	1.000	1.000	0.996
355487	0.996	0.996	0.996	1.000

P-value

	GP04.071927035	355511	355510	355487
GP04.071927035	1	0	0	0
355511	0	1	0	0

355510	0	0	1	0
355487	0	0	0	1

Haplotype by GYNGroup

PHASE Haplotypes (All)

	Case	Case(%)	Case.X^2	Control	Control(%)	Control.X^2	OR	ln.OR
TTGC	0	0.00	0.48	1	0.10	0.44	0.0000	-Inf
CCAC	262	25.99	1.03	319	31.65	0.95	0.7586	-0.2763
TCGT	1	0.10	0.00	1	0.10	0.00	1.0000	0.0000
TTGT	220	21.83	1.41	203	20.14	1.30	1.1071	0.1017
CCAT	1	0.10	0.56	0	0.00	0.52	Inf	Inf

Pearson Chi-squared Test = 6.6985, DF = 4, P-value = 0.1527

Permutation Test P-value = 0.56

PHASE Haplotypes (Low Frequency Excluded)

	Case	Case(%)	Case.X^2	Control	Control(%)	Control.X^2	OR	ln.OR
CCAC	262	26.10	1.03	319	31.77	0.95	0.7582	-0.2768
TTGT	220	21.91	1.41	203	20.22	1.30	1.1072	0.1018

Pearson Chi-squared Test = 4.4162, DF = 1, P-value = 0.0356

haplo.score Haplotypes

	Hap.Freq	Score	P.X^2	P.Sim
CCAC	0.5772	-2.3513	0.0187	0.0168
TTGT	0.4208	2.2111	0.0270	0.0249

Global Score = 7.5085, DF = 2, Global P.X^2 = 0.0234, Global P.Sim = 0.0117

[0285] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

[0286] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the claims which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

What is claimed is:

1. A method for identifying a subject at risk of breast cancer, which comprises detecting the presence or absence of one or more polymorphic variations associated with breast cancer in a nucleic acid sample from a subject, wherein the one or more polymorphic variations are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby the presence of the polymorphic variation is indicative of the subject being at risk of breast cancer.

2. The method of claim 1, which further comprises obtaining the nucleic acid sample from the subject.

3. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 1 selected from the group consisting of 133, 7938, 8873, 13221, 17288, 25732, 26923, 39977, 41284, 41410, 41477, 41514, 42606, 42742, 59515, 59808, 60265, 67152, 68332, 71128 and 76427.

4. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 1 selected from the group consisting of 7938, 26923, 39977 and 59808.

5. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in a region spanning positions 7938-59808 in SEQ ID NO: 1.

6. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 2 selected from the group consisting of 201, 6395, 8558, 9429, 9809, 10072, 10511, 11556, 16857, 16951, 17027, 17177, 17615, 17950, 18329, 18384, 18561, 18579, 18871, 27152, 27306, 28091, 28661, 29011, 29962, 29969, 30085, 31656, 31685, 31749, 45389, 45459, 46647, 49860, 53061, 57308, 61563, 61660, 62212, 67090, 67198, 70071, 70191, 74006, 75600, 85761, 90798, 90883, 91259, 95416, 95446, 96368, 97050, 97362, 97630, 97989 and 98107.

7. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 2 selected from the group consisting of 10511, 11556, 17177, 18384, 28661, 31656, 31685, 31749, 45389, 45459, 46647, 49860, 53061, 57308, 61563, 61660, 67090, 67198, 70071, 74006, 75600, 85761, 90798, 90883, 91259, 95416, 95446, 96368, 97362, 97630, 97989 and 98107.

8. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in a region spanning positions 10511-98107 in SEQ ID NO: 2.

9. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 3 selected from the group consisting of 160, 6053, 9719, 10481, 10676, 17179, 18561, 18658, 18694, 18858, 24582, 24683, 24767, 27402, 28150, 28494, 32003, 35588, 35619, 35856, 36254, 37314, 40033, 40095, 42593, 42799, 43090, 46683, 49774, 51796, 52079, 53857, 53971, 55899, 60682, 61291, 72720, 72752, 85507 and 89751.

10. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 3 selected from the group consisting of 160, 6053, 18658, 18694, 18858, 24683, 27402, 28494, 32003, 35588, 35856, 40095, 46683, 52079, 53857, 72720 and 72752.

11. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in a region spanning positions 160-72752 in SEQ ID NO: 3.

12. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 4 selected from the group consisting of 196, 13311, 14486, 14691, 15551, 17702, 17872, 19588, 19910, 20006, 20575, 21092, 22830, 23455, 23716, 23890, 24001, 24995, 27282, 27779, 29099, 31185, 33994, 34942, 35137, 36538, 37139, 37358, 38828, 39469, 40233, 40472, 41679, 41682, 42831, 42976, 44128, 44195, 46769, 47363, 48843, 52574, 52602, 53212, 53781, 54710, 55808, 57987, 58556, 59148, 59286, 60217, 60412, 60753, 60791, 61524, 62543, 62825, 62826, 62857, 63400, 63960, 64307, 64539, 65728, 66000, 66521, 68185, 69643, 74909, 82973, 83039, 85713, 86873, 90293, 91810, 92609, 92884 and 42831.

13. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 4 selected from the group consisting of 196, 13311, 14486, 19910, 20575, 23716, 23890, 24995, 29099, 33994, 34942, 37139, 40233, 40472,42831, 42976, 44195, 48843, 58556, 59286, 60217, 62826, 62857, 63400, 63960 and 74909.

14. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in a region spanning positions 196-74909 in SEQ ID NO: 4.

15. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in linkage disequilibrium with one or more positions in claim 3, 6, 9 or 12.

16. The method of claim 1, wherein detecting the presence or absence of the one or more polymorphic variations comprises:

hybridizing an oligonucleotide to the nucleic acid sample, wherein the oligonucleotide is complementary to a nucleotide sequence in the nucleic acid and hybridizes to a region adjacent to the polymorphic variation;

extending the oligonucleotide in the presence of one or more nucleotides, yielding extension products; and

detecting the presence or absence of a polymorphic variation in the extension products.

17. The method of claim 1, wherein the subject is a human.

18. A method for identifying a polymorphic variation associated with breast cancer proximal to an incident polymorphic variation associated with breast cancer, which comprises:

identifying a polymorphic variation proximal to the incident polymorphic variation associated with breast cancer, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic variation;

determining the presence or absence of an association of the proximal polymorphic variant with breast cancer.

19. The method of claim 18, wherein the incident polymorphic variation is at a position in claim 3, 6, 9 or 12.

20. The method of claim 18, wherein the proximal polymorphic variation is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the incident polymorphic variation.

21. The method of claim 18, which further comprises determining whether the proximal polymorphic variation is in linkage disequilibrium with the incident polymorphic variation.

22. The method of claim 18, which further comprises identifying a second polymorphic variation proximal to the identified proximal polymorphic variation associated with breast cancer and determining if the second proximal polymorphic variation is associated with breast cancer.

23. The method of claim 22, wherein the second proximal polymorphic variant is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the proximal polymorphic variation associated with breast cancer.

24. An isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(d) a fragment of a nucleotide sequence of (a), (b), or (c); and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

wherein the nucleotide sequence comprises one or more polymorphic variants associated with breast cancer selected from the group consisting of a thymine at position 7938 in SEQ ID NO: 1, a cytosine at position 26923 in SEQ ID NO: 1, a thymine at position 39977 in SEQ ID NO: 1, a thymine at position 59808 in SEQ ID NO: 1, a thymine at position 10511 in SEQ ID NO: 2, a cytosine at position 11556 in SEQ ID NO: 2, a thymine at position 17177 in SEQ ID NO: 2, a thymine at position 18384 in SEQ ID NO: 2, an adenine at position 28661 in SEQ ID NO: 2, an adenine at position 31656 in SEQ ID NO: 2, an adenine at position 31685 in SEQ ID NO: 2, a guanine at position 31749 in SEQ ID NO: 2, a thymine at position 45389 in SEQ ID NO: 2, a guanine at position 45459 in SEQ ID NO: 2, an adenine at position 46647 in SEQ ID NO: 2, a thymine at position 49860 in SEQ ID NO: 2, a thymine at position 53061 in SEQ ID NO: 2, an adenine at position 57308 in SEQ ID NO: 2, a guanine at position 61563 in SEQ ID NO: 2, a guanine at position 61660 in SEQ ID NO: 2, a guanine at position 67090 in SEQ ID NO: 2, a cytosine at position 67198

in SEQ ID NO: 2, an adenine at position 70071 in SEQ ID NO: 2, a cytosine at position 74006 in SEQ ID NO: 2, an adenine at position 75600 in SEQ ID NO: 2, a guanine at position 85761 in SEQ ID NO: 2, a thymine at poisition 90798 in SEQ ID NO: 2, a cytosine at position 90883 in SEQ ID NO: 2, an adenine at position 91259 in SEQ ID NO: 2, a cytosine at position 95416 in SEQ ID NO: 2, a thymine at position 95446 in SEQ ID NO: 2, a thymine at position 96368 in SEQ ID NO: 2, a thymine at position 97362 in SEQ ID NO: 2, an adenine at position 97630 in SEQ ID NO: 2, a cytosine at position 97989 in SEQ ID NO: 2, a thymine at position 98107 in SEQ ID NO: 2, an adenine at position 160 in SEQ ID NO: 3, a guanine at position 6053 in SEQ ID NO: 3, a guanine at position 18658 in SEQ ID NO: 3, a guanine at position 18694 in SEQ ID NO: 3, a thymine at position 18858 in SEQ ID NO: 3, a guanine at position 24683 in SEQ ID NO: 3, a guanine at position 27402 in SEQ ID NO: 3, a thymine at position 28494 in SEQ ID NO: 3, an adenine at position 32003 in SEQ ID NO: 3, a cytosine at position 35588 in SEQ ID NO: 3, an adenine at position 35856 in SEQ ID NO: 3, a guanine at position 40095 in SEQ ID NO: 3, an adenine at position 46683 in SEQ ID NO: 3, an adenine at position 52079 in SEQ ID NO: 3, a cytosine at position 53857 in SEQ ID NO: 3, an adenine at position 72720 in SEQ ID NO: 3 a cytosine at position 72752 in SEQ ID NO: 3, an adenine at position 196 in SEQ ID NO: 4, a guanine at position 13311 in SEQ ID NO: 4, a thymine at position 14486 in SEQ ID NO: 4, a thymine at position 19910 in SEQ ID NO: 4, an adenine at position 20575 in SEQ ID NO: 4, a guanine at position 23716 in SEQ ID NO: 4, a guanine at position 23890 in SEQ ID NO: 4, an adenine at position 24995 in SEQ ID NO: 4, a cytosine at position 29099 in SEQ ID NO: 4, a thymine at position 33994 in SEQ ID NO: 4, a thymine at position 34942 in SEQ ID NO: 4, a thymine at position 37139 in SEQ ID NO: 4, a thymine at position 40233 in SEQ ID NO: 4, an adenine at position 40472 in SEQ ID NO: 4, a guanine at position 42831 in SEQ ID NO: 4, a guanine at position 42976 in SEQ ID NO: 4, a thymine at position 44195 in SEQ ID NO: 4, a thymine at position 48843 in SEQ ID NO: 4, an adenine at position 58556 in SEQ ID NO: 4, a guanine at position 59286 in SEQ ID NO: 4, an adenine at position 60217 in SEQ ID NO: 4, a cytosine at position 62826 in SEQ ID NO: 4, a thymine at position 62857 in SEQ ID NO: 4, a thymine at position 63400 in SEQ ID NO: 4, an adenine at position 63960 in SEQ ID NO: 4 and a cytosine at position 74909 in SEQ ID NO: 4.

25. An oligonucleotide comprising a nucleotide sequence complementary to a portion of the nucleotide sequence of (a), (b), (c), or (d) in claim 24, wherein the 3' end of the oligonucleotide is adjacent to a polymorphic variation associated with breast cancer.

26. A microarray comprising an isolated nucleic acid of claim 24 linked to a solid support.

27. An isolated polypeptide encoded by the isolated nucleic acid sequence of claim 24.

28. A method for identifying a candidate molecule that modulates cell proliferation, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence in SEQ ID NO: 1-4;

(ii) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(iv) a fragment of a nucleotide sequence of (i), (ii), or (iii); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate molecule that modulates cell proliferation.

29. The method of claim 28, wherein the system is an animal.

30. The method of claim 28, wherein the system is a cell.

31. The method of claim 28, wherein the nucleotide sequence comprises one or more polymorphic variations associated with breast cancer.

32. The method of claim 28, wherein the one or more polymorphic variations associated with breast cancer are at one or more positions in claim 3, 6, 9 or 12.

33. A method for treating breast cancer in a subject, which comprises administering a candidate molecule identified by the method of claim 28 to a subject in need thereof, whereby the candidate molecule treats breast cancer in the subject.

34. A method for identifying a candidate therapeutic for treating breast cancer, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence in SEQ ID NO: 1-4;

(ii) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(iv) a fragment of a nucleotide sequence of (i), (ii), or (iii); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for treating breast cancer.

35. The method of claim 34, wherein the test molecule inhibits cell proliferation or cell metastasis.

36. A method for treating breast cancer in a subject, which comprises contacting one or more cells of a subject in need thereof with a nucleic acid, wherein the nucleic acid comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(d) a fragment of a nucleotide sequence of (a), (b), or (c); and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

whereby contacting the one or more cells of the subject with the nucleic acid treats breast cancer in the subject.

37. The method of claim 36, wherein the nucleic acid is RNA or PNA.

38. The method of claim 37, wherein the nucleic acid is duplex RNA.

39. A method for treating breast cancer in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with breast cancer in a nucleic acid sample from a subject, wherein the one or more polymorphic variation are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic variation; and

administering a breast cancer treatment to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

40. The method of claim 39, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 6, 9 or 12.

41. The method of claim 39, wherein the breast cancer treatment comprises a nucleic acid comprising a nucleotide sequence complementary to a nucleotide sequence in SEQ ID NO: 1-4.

42. The method of claim 41, wherein the nucleic acid is a double stranded RNA.

43. The method of claim 39, which further comprises extracting and analyzing a tissue biopsy sample from the subject.

44. The method of claim 43, wherein the treatment is chemotherapy, surgery, radiation therapy, and combinations of the foregoing.

45. The method of claim 44, wherein the chemotherapy is selected from the group consisting of cyclophosphamide (Cytoxan), methotrexate (Amethopterin, Mexate, Folex), fluorouracil (Fluorouracil, 5-Fu, Adrucil), cyclophosphamide, doxorubicin (Adriamycin), and combinations of the foregoing.

46. The method of claim 45, wherein the combinations are selected from the group consisting of cyclophosphamide (Cytoxan), methotrexate (Amethopterin, Mexate, Folex), and fluorouracil (Fluorouracil, 5-Fu, Adrucil); cyclophosphamide, doxorubicin (Adriamycin), and fluorouracil; and doxorubicin and cyclophosphamide.

47. The method of claim 39, wherein the breast cancer treatment reduces breast cancer metastasis.

48. A method for detecting or preventing breast cancer in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with breast cancer in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic variation; and

administering a breast cancer prevention procedure or detection procedure to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

49. The method of claim 48, wherein the one or more polymorphic variations are detected at one or more positions in wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 6, 9 or 12.

50. The method of claim 48, wherein the breast cancer detection procedure is selected from the group consisting of a mammography, an early mammography program, a frequent mammography program, a biopsy procedure, a breast biopsy and biopsy from another tissue, a breast ultrasound and optionally ultrasound analysis of another tissue, breast magnetic resonance imaging (MRI) and optionally MRI analysis of another tissue, electrical impedance (T-scan) analysis of breast and optionally of another tissue, ductal lavage, nuclear medicine analysis (e.g., scintimammography), *BRCA1* and/or *BRCA2* sequence analysis results, thermal imaging of the breast and optionally of another tissue, and a combination of the foregoing.

51. The method of claim 48, wherein the breast cancer prevention procedure is selected from the group consisting of one or more selective hormone receptor modulators, one or more compositions that prevent production of hormones, one or more hormonal treatments, one or more biologic response modifiers, surgery, and drugs that delay or halt metastasis.

52. The method of claim 51, wherein the selective hormone receptor modulator is selected from the group consisting of tamoxifen, reloxifene, and toremifene; the composition that prevents production of hormones is an aramotase inhibitor selected from the group consisting of exemestane, letrozole, anastrozol, groserelin, and megestrol; the hormonal treatment is selected from

the group consisting of goserelin acetate and fulvestrant; the biologic response modifier is an antibody that specifically binds herceptin/HER2; the surgery is selected from the group consisting of lumpectomy and mastectomy; and the drug that delays or halts metastasis is pamidronate disodium.

53. A method of targeting information for preventing or treating breast cancer to a subject in need thereof, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with breast cancer in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic variation; and

directing information for preventing or treating breast cancer to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

54. The method of claim 53, wherein the one or more polymorphic variations are detected at one or more positions in wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 6, 9 or 12.

55. The method of claim 53, wherein the information comprises a description of a breast cancer detection procedure, a chemotherapeutic treatment, a surgical treatment, a radiation treatment, a preventative treatment of breast cancer, and combinations of the foregoing.

56. A method of selecting a subject that will respond to a treatment of breast cancer, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with breast cancer in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO: 1-4;

(b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4 ;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-4 ; and

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic variation; and

selecting a subject that will respond to the breast cancer treatment based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

57. The method of claim 56, wherein the one or more polymorphic variations are at one or more positions in wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 6, 9 or 12.

58. A composition comprising a breast cancer cell and an antibody that specifically binds to a protein, polypeptide or peptide encoded by a nucleotide sequence identical to or 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-8.

59. The composition of claim 58, wherein the antibody specifically binds to an epitope that comprises a glutamine at amino acid position 278 in SEQ ID NO: 9 or a glycine at amino acid position 389 in SEQ ID NO: 12.

60. A composition comprising a breast cancer cell and a RNA, DNA, PNA or ribozyme molecule comprising a nucleotide sequence identical to or 90% or more identical to a portion of a nucleotide sequence in SEQ ID NO: 1-8.

61. The composition of claim 60, wherein the RNA molecule is a short inhibitory RNA molecule.

FIGURE 1-A

>3:198232901-198309500

```
1       actaagttac tacaaagcag ttaacatagt ctcagatatt aaaaatttaa gatactgaag
61      agctaagtca attcctgata atttcttaaa tttgtgatct atttctttca ctgtattcag
121     atcttcatat ttRtagtctt tgcttagagt cttcccaccc gcccccacct cgctctgttg
181     ctcaggccgg agtgcagtgg tgcaattccg gctcatcgca gcttctgcct cctgggttca
241     agtgattctt gtgcctcagc ctcctcagta gctgaaatta caggtatgta ccaccatgcc
301     tgtttaattt ttatgttagt agagacagag tttcatcaag ttggccaggc tggtcttgaa
361     ctcctgcttt caagggattt gcctgctttg gcctcccaaa gtgctgggat tacaggcgtg
421     agccattgcg cctggccaag tctttgcctt actcttaaca attcctcaca gctacctttc
481     tatatgttca aaacatactt cggcatttta aagttacttt tcctttaaac cttctatttc
541     atttacctac ttaagattcc taccttttcc ttttgtttca atcaattgat acttttagtc
601     ctatcagtct ccacatttta cctatcctgt gtatattttc gtaaatactc ccaccaacat
661     atttttgttt tgtttttaag gttttttttgg agacaggctc ttgctctgtc atccaggctg
721     gagtgcagtg gcacgatcat ggctcatagt agctctgacc tcctgggctc aagcaatctt
781     cctaccttag cctctcaagt agctgagcct acaggcatgc atcactatgc ccagctcatt
841     tttaaaatta cttgcagcga cagggtgtcc tatgttaccc aggggatctt caactcttag
901     gctcaagtga tcctcttgtc tcagcctccc gaagtgctaa ccactacacc tattttctta
961     aatttttatt ttgacggaag agtttcatcc ctttttttcag ttaagtcatt tacactcctc
1021    tcctagaatt acatcaaaac tcagcaaagc tataatttta cataagagat tttctccatt
1081    caacatctca aaaattcctg tatcaaacca tatttcttgt catgaaacgt attaggacgt
1141    atctgaactg ctggctacaa gttagtatgt atacgcttat acaaggagag atatttaaaa
1201    gctgattttt tcatttgata tccccctttca aattttaacc tttgttttca ccgccacact
1261    agaagtaaaa agtgaggaag gggtatgggg gaaattggac ttcctcaatc tcttcttaac
1321    atgcaatccc tataaagaaa aacattcatt tataactgta ataaaatata ttaaacacag
1381    atattaaacc caagagaaac aattctcaga ggatggatac tattaattgg cctaaagaaa
1441    aagggaaact atttttaaga aggccagtta gttgttttaa gtacaatagc tctctggcat
1501    actcggagga ttagttccag aaccccacgt ataccaaaac ccgtgcatgt tcaagtctaa
1561    caggaggcct tgtggaatct gacaggcata tggcattttc tatcccagtt tggttgaaca
1621    gcacgtgcgt aagtggacac tcgcagttca aacctgtgtt gttcaagggt caactgcagc
1681    acaggcgcgt agaaaaatac agcaagggaa gtgggaaagc aggagactga aaaaaatttt
1741    aattacctat ctactactga aaatgataag aaatgacaaa tgtctgcaaa taaatgccac
1801    tagtccactg caatcttccc agattatctc cacctcttct ccctttggtt ggtataagaa
1861    gtaataatga aaataaagtc aggacactgg gtgttagaaa tctacctttg taactcaaga
1921    tttcaatcaa agctttgaag cgcagaaaag ttaagtgact atggcctaac ctctctagac
1981    tctggtcact tgcaaaatag ggattatgcc accaaacagt ttttttcttg cgaaaataaa
2041    atatgtcaca aatataagac ctaaccaatg ggaaagagaa aataaattta acttatatat
2101    tgagtttctt tccaatttat ttcttttatg gtattattac aatataatta atgctttctg
2161    aacaaaagat tgagatataa gctctgaaag ttggcctaat actcttgagg aacaataaga
2221    ttaaatattt gaggtgagag attccaagag tttaatgagt agttgaaaga gtaagaacag
2281    ggtatggaga cccacccaga accccagagt tgccacaaag attttaagct gaagatatct
2341    gagattcaac agatgcagga aaaaagcctt ctcggagttt ctctaatctg actaaaaggt
2401    gcagcttccg agaaatgaag ttaccataaa taccctctga ggagtctgat ggccccgaag
2461    aatggaaaga ccagtcatac cagtttagac aaatgtcatc acaaacttta tcccccatct
2521    cttctcctaa aaacttgtct ttcctgaaga aatgtacgtt tttctaataa aagctgtttc
2581    tccctccttc ctttccctac tagattaggt gtacaagctt ttaacttcat ccacaagcta
2641    gcttcagtat ttcttctgtt gcctcttgca tccatataca agaaactttt tttccatgtt
2701    aatctgtttt ttttgtcagt ttacttcaca ggcccgttac tgaacataca agggtggagg
2761    aaaaggtttt tcttcgctta aaaggggcta gaagaaatct taaccgttta acagttgtgt
2821    aactagcaaa ttacttggtc tctttatgga agagcttatt caactacaac gtaaaacggg
2881    gataacggtg cttacttaca ggattatctt gggagttaaa agagtgaata aacataacat
2941    ggttcgtatc atgcatggcg tacagtttta ctattatata ctatcacagg agattctttt
3001    ttttttttttg agacagagtc tcgctctgtc gcccaggctg gagtgcagtg gcgtgatctc
3061    ggctcactgc aagctcggcc tcccgggttc acgccattct cctgcctcag cctcgcgagc
3121    agctgggact acaggtgccc gccaccacgc ctggctaatt ttggttttgt atttttagta
3181    gagacggggt ttcaccgtgt tagccagaat ggtctctatc tcctgacctc gtgatctgcc
3241    cgcctaggcc tcccaaaatg ctgggattac aggcgtgagc caccgcgcct ggccaagaga
3301    ttcttaaact ataggaagta tgtgataaaa ctttataact ttgggatatg gagtccaaaa
3361    cgagtatagt attatatttc cagttcccat gttcgattat ttttatttttt attttgagac
3421    ggggtctcac tctgtcgccc aggctcaagt gcagtggtga gatcatagca atcctgcctt
3481    ggctgggact acgggtgcgc actaccacgc atggctaatt aaaaaacttt aaaaaaagtt
3541    ttagcagaga caagatcttg ctatgttgcc caggctggtc tcgaactcct gggctcaagt
3601    aatcctccca cctaggcctt ccaaagtgct gggattagag gcctgagcca ccacgtccag
3661    ctaatttctg ataatgattc tccaaaaaga ttaacagtaa aggcccccca aagtgaatta
```

FIGURE 1-B

```
3721  ttccttgcct ttggaattaa taaaactgaa aggagttaga agataatgtt ccttcatatg
3781  tcttgtagag attatttttt ctcataatca gaatccaagt caagaaagtt caaagtacac
3841  aagctgaaaa acccataaaa gtacttttt ggaagactaa taaggcaaac catagatcct
3901  gaaagataaa aaacttaatt ctaccaaaag cttacctttg tctctatctg tctgtcttgt
3961  atgtatatat ataagattat caacaaggca gagaatccca agcctttagg aaagatacta
4021  actgcaagat gctgattatc agagtgtcgg atgactcctt ttgtacttct taaaatccga
4081  tttgtgttaa tgtttttgtg ttgtgtgtgc gtgtagctaa cctgttacta agctttgatg
4141  cttttcccca cctcttatta tctttgtatg aatgtgaggc tacttttaaa ttacttatta
4201  ctggggctgg gggctgtggc tcacgcctgt aatcccagca ctttgggagg ccgaggcagg
4261  cagatcacct gaggtcagga gaccagtctg gccaacatgg tgaaaccccg tcgctattaa
4321  aaatacaaaa ataagccggg cgtggtggca ggtgcctgta atcccaccta cttgggaggc
4381  ttgaacctgg gaggcggagg ttacagtgag ccgagattgt gccactgcac tccagcctgg
4441  gcaagagagt gaaactccat cttaaaacaa acaaaaaaaa ttacttatta ctttgctaga
4501  ttttgctaga gtagctattc cagaaagtcg aggggtgcag gaaagataga ggcacaggac
4561  atttttacag gttttaaaac ccaaatactg tcacattatc tgatcctaca gaggcaggct
4621  gctttatcca aatacgggtc ctctgctttt ctaaaccaaa ccacattatg agtgaaattt
4681  tataccatat cctaaatttg tgccatcaca aacattctca acattccata ttttaactct
4741  tcccctcttt ccgtgtggct tctgcaagac ctcagctgct taaggagtct ttacctctac
4801  ttggttgtcc aggtatcata cctgccttct tagttttagt taatattaag tgaagttgtg
4861  tgggtaggag catgcgtgta tgttttgttt ttcttactgc ttttgtgtgt tctgtgtgtt
4921  ttgaggttag ggagcagata ttaagttacc cgtaaccata cctaaaaagt ttatattcct
4981  tataaacata taaatgagag gctgtctata atactaagga accagtgact gtaactcggc
5041  ttccatatag tgtacctacg gggtggaata tgttggccag aagataaaag tacaacatag
5101  ggtcccattg gcagatttgt ttctcacaaa aattaacgaa ttgacattct gacttcatac
5161  gttagtctct attgtttggt taatgtgaat taatattcag aatttactcc caattaaaca
5221  aattttcccc tgtttagtta ctaagcctaa aaggccaagg ttttactaga ttgatagaac
5281  ctcccatact taagaaagac caagcaatgc tcaacatgct gctttgatca gctaaaaaac
5341  atactagatt aaaaaagaaa ttctgcttca aaaaagaata aatgtaatta gttgataaat
5401  tacagaatta ctcatttagc tcctttactc acatcatctt ttccttattc atcatgttgt
5461  ttatgagcat caacaaatat gctatgtatt cttttgggta tttaagcttt ctaaattcca
5521  ctgaggtttt attggagatg tggttatgct gatgtacact agagtcttac atatgtttac
5581  tatggcatga cctcaaattc tccatcacca ccttcttccc ccaggaactt agtaaagtca
5641  catccaccat acttcaccat ctagacacac agtatactcc aaacttggca ttcttctctt
5701  tattttaaat agcggcagta acagctatta gaaacatgtc ttaaatcttg atattaataa
5761  ttataaccta attcaaaagc tagaaaaaat ctgggtgacc taatttttt gggcattatt
5821  actatatctg cttatggttt ttattatttt ttcagatcaa taaagtagta aaagggacat
5881  taaaagctat gtcttcagtt acagaacaga tttaaaaaag taaacagtta aattaaccca
5941  cataatgggg gttttgtggg ttaagcttgt ggatgccagt tcccacatca ttaaaaaata
6001  cttttggcag ggtgcagtgg ctcacaccgg taatcacagc actttgggag gctgaggtgg
6061  gcggatcgcc tgaggtcagg agtttgagac cagcctgacc aacatggaga aactctatct
6121  ctactaaaaa tacaaaaaat tagccggtat ggtggtgcat gcctgtaatc ccagttactt
6181  gggaggctga ggcaggagaa tcacttgaac ctgggaggca gaggttgcgg tgagctgagt
6241  tcgcaccact gcactccagc ctgggcagta agagtggaac cccgtctcaa caaacaaaaa
6301  acaaaaacca aaaaaaacca aaaaaacaaa caacttttga tctggaaaaa gaactaggtt
6361  cagactaaac aaatactgga ttagcaaatc agaaaaaaag cacaaaaaag gaagtttagc
6421  ttggaagtta aacattatta aatattaaag aggaagaatt ttaaaaacta aaataagcaa
6481  taactataga caataagaat gttattacct tcaggtcgat attgtgcaac aattgtgaca
6541  gcctggccag cattttcaa tgcagctgct gcctgctcat gactagcagc tctgaggtca
6601  acactgttta cctgtaaatc aaaccaaatc ttaatttggg agaaaagaat cactgtgatt
6661  agaaatcaca atagtctatt ctttgagtaa gcatgtatta aaaaattagt tgcaataact
6721  ttaccactaa aatgtttctt gttctgattg actaaaaata tattgtgtag taaaacaaaa
6781  tattatcatt caaagctatt ttatttacta aaaccacctt gcttaatgtc aagttagtaa
6841  aaccaaaaca ttttccccta gaaggctggg tatcaatcca ttgagaaatg taccttaatt
6901  tgagaacctg aaaaatcaag tggcatcatt ttgcacactt tattttgcta ggaactcaag
6961  caatgggagg gaaagtcatt ttgcttgtag catccattta cctcctgcca cattctcacg
7021  cctactgctc aattcagttc caattatttt tcttcctgta caaactaaaa cttccattct
7081  tttggtataa taggaaagct attatttaga aatattaaaa cagctcagat ttcaaggact
7141  gatttaatta gaataaatgt taattaaatg aaccaggagt tatttttga attaagagcc
7201  cttccttagg tgaaataatt atccaattct accagtaggt acaatcctac tccagaagaa
7261  gaggcaactt aattcttatt tctcaaaggt taccaatact ttgcgcgcat atatacatat
7321  atagtgcaag gtacattatc ttaagagatt gttccttttc cacaataaga caaacttaag
7381  aggacattca tggaagcatt atacttgaat gtaataaact ttaaataccg aattaaaaaa
7441  aattaagaag caacaatgaa tagaatatgg caatgttcag gaccttcact gaaaattttt
7501  ggaaaggaac gatcatcatg caatttttac acaaaaactt gttttgttga gaactaatct
```

FIGURE 1-C

```
7561   aaaatctgtt aaagatcaca tatcactttt aaaaaatttg actatgagcc cagaaattct
7621   tgggatcata aatttacata aagaacattt taaccatata aagatctgat gttttatatt
7681   atctaaaaag aaaccattat gcaactaaat ttaaatacat acccccttgag gctgggcatg
7741   gtggctcaca cctgttatcc cagcactttg ggaagctgag gtgggtggtt cacttgaggc
7801   caggagtctc agaccagcct ggccaacatg gcaaacacct gtctctacta aaaatacaaa
7861   aattagccgg gcatggtggt gtgtgcctgt agtcccagct actcaggagg ctgaggcatg
7921   agaatcactt gaacctgRga ggcggaagtt acagtgagct gagatcacat cactgtactc
7981   cagcctgggt gacagagaga gggacgctct gtctccagcc agccccccgc acccccctca
8041   atgaaaacca agatctaaaa tccataacat aaacaaatgt aaacaatgtg gttattagta
8101   tctccagctg atgatggcca tataaaactc aatacaatgt ataaacagtt ctacttagtg
8161   cataaagaaa atcctcccaa ttcctgggag aacatatcag gcaagaaata taaagaaaga
8221   caggtttttc acagattatt gtagaaaaga ttaagaatgt taatgaatta tttctgtgca
8281   ccaaactctt ttaagagcat aatttgttga acaaaactac atattaaaga ctttccagac
8341   tatcactgaa catttcaatt gaagatacat tattaaggat attttagaag aaattccagt
8401   actgagagag atacttgtat aaatagcctt agagtcactt tttttctttt acaccgacat
8461   ttttaaaaat ggagaaataa ttcgcatacc atagaattca catttttaaa gcatatagtt
8521   tagtggtttt tggtatattt gcaagtctgt gcaaccatca ttaccaccta attccaaaat
8581   attttcatta ccccaaaaag aaactctgta tccactgaca gtcactcctg tgtctcctct
8641   ccatccctgg caactacgaa gctactttct gaccatactg tacatatacc tatatatacc
8701   tattcggtat atttcaaata aacgggatat ataatatgta gccatctgtg cttggcttct
8761   ttcacttatt ttccaggttc acccttggag catgtatcag tacttcattc cttatattta
8821   gtggcgctaa ataatgcgcc actaaaatcc attaaaaatg gatatactgg ccRggtgtgg
8881   tggctcacgc ctgtaatccc agcactttgg gaggccaagg caggcagatc acgaggtcag
8941   gagatcgaga ccttcctggg taacatggtg aaaccctgtc tctactaaaa atacaaaaaa
9001   attagccggg cgtggtggca ggcacctgta gtcccagcta cttgggaggc tgaggcagga
9061   gaatggcgtg aagctgggag gcggagcttg cagtgaacag agattgcgcc actgcactcc
9121   agcctgggcg acagagcaag actctgtctc aagaacaaaa aacaaaaaaa ccaaacccaa
9181   aaaaaaaagg atataccact tttgttgatc catgtcctta aatttatact taagtatttt
9241   attctttttg atagtattat aaatcaatgt tttctttgtc tatcctgcaa atttaactga
9301   actcatttat tagcttgaca gtctgttgtt gtatttctta agatttttтc tatacaagat
9361   tatgtcatct gcaaacagaa atagttttct tgctttttat tctggtagcc ttttttttt
9421   tttaatcttg cctaatttct agaacctcca ggacaatgtt gaatagaagt aggaagagca
9481   gataccctg tcttgttcct gattttagag ggaaagcttt cagtctctca ccatttagta
9541   tgatgtcagc tatagggttt ccagagatac cctttaacag gttgaagttc ccttttattc
9601   ctagtctatt tagtgttctt atcatgaaaa gctgttggtt tatgacaatt gctttttctg
9661   tgtctactga gatatcatgt gctttatgtc ctttattaag atggtgtctt actttgattg
9721   gtttttctat gttgaactaa ccatgcattc cttgaataaa ttccatttgg ttatggtaca
9781   tcatccttgt tcatgctgtt ggatccaggt tgttattttg ttgattttta catctatatt
9841   catatgggat attggtttgt aatttttttct tcttgtgatg tctttggttt ggtataagga
9901   taatactggc cacacagaat gagctggaaa gcattcctca tgcttattcc tttttgggga
9961   agagtttgtg aatgattggt gttaattctt taaatgtttg gcagaatgta ccagtgcagc
10021  catctggtcc tgggattttt gtgttttgtt agttttgat tactaattca atttccttgt
10081  tttaaatcta ttaaaattct ctatttcttc ttgagttagt ttcagtagtt tatgtctttc
10141  tagaaatgta ttggtaatta gagtatataa ctataaaata tataccatct atgactacat
10201  aatgatatac tataaatgtc tttcgccttt agtaataatg tttttcttca agcgtccttt
10261  gatattagtc atattagtac agccacagta ggtgtctttt ggttactttt tgtgtaatgc
10321  atcttttcca cccgtttact ttcaaccttt ttgtgtcttt gagtctaaag tgtctctctt
10381  gtagaccctg tgtagttgga ttctgttttc taatctactt tatcaacatc tttggccttt
10441  taactggagt gcttaatctg tttatattta atataattac tgataatgta gaatttactt
10501  aatttggtta tttgttctat gtcttttctt tctctattcc tcaattactg cattattttt
10561  atattaaata gatatttcct agtgtaccat tttaattccc atgtcattaa ttttactaca
10621  tatatttgtt agtcattttc ttagtggtta tcttggatta taagtaaaaa tttaaagcca
10681  gatgtggtgg ttctagactg taatcccagt gactcaggtg gctcaggcaa aaggactgaa
10741  gcaggagcat cacttgagta caacactttg agggctggct aggcatcaca gtgagaccct
10801  tatctctaaa aaaataaaga aaacaattca gccaggtgta gtggtgcata cctgcagccc
10861  agctactcgg gaggctgagg cacgaggatt gcctgagtaa gtttgaggct gcagtgagct
10921  atgactgtgc tactgcactc cagcctgggg gacagagtga gatgctgcct ctaaaaataa
10981  ataagtaaac aaacaaacat tttataagaa cctagtttct ttacccgcct tgtggtgttg
11041  tcatattaat tttatgttta tacgatgtgt gcacatcaaa tttgtaatta ttgctttatt
11101  cagttgtttt taaaatcagg ttggagaaaa aaagagttgc atataaaaca ataccttat
11161  gttgttttat atttacctgt gtagttgcct ttacttaggc tctttatttc ttcatgtgga
11221  tttgagttac tctctagttt cctttcattt tagtctgaag gattctctta tttcctatag
11281  ggcaggtttt tgagagatga actccttcag cttttgttta tctggcaatg tcctgatttc
11341  ttctttattt tcaaaagata gttttgctgg atatagagtt ctagtcaaca aagttttttt
```

FIGURE 1-D

```
11401  cttttcagaa ttttgaattt gttatttcac tgctttctga cctccatggt ttctgatgag
11461  aaatcagcta ttaaccttgt tgaggatccc ttgcacgtgg tgagtcattt ctttttgcta
11521  ctttcaatat tctgccttcg gctttagaca gttcgattgt aatgtgtcta tgttgtagat
11581  ctctttaaat taatttaact tggagttcac tgaggttctt ggattaatgg ttttaatcaa
11641  gttctagaag tttttgggcg agtatttatt caaatattct ttcttctttt ttctctcttt
11701  cccatctttc acatactccc attatgttgg taaatttgat agtatcccat aggtctctga
11761  agctgttaat ttttcttcat tctttctttc tttgttattg ttcagttctt cagcatcaaa
11821  atctcaattg acccatcttc aagtttgctg attctttctt ctgccagctc aaatctgctg
11881  ttgaccttgt ctagagaatt tttcatttaa gttactgtac tttccaactc cagaatttct
11941  atttggttct tacatataat ttctaactct ttattgatat tttttacatg gtgagatatt
12001  gttccttttac tttcatgtca gtctacagga tggttttctt ttattccttg aacgtattta
12061  aaatacctgg tttaatgtct ttgtttagca tgttcaatgt ctgaccttcc ctaggaacga
12121  tatccattga ttgcgttttt ctcctgtata tgggccatac ttttctttcc gcttgtctca
12181  aaattttatg acgaaaaata ggcatttaaa ataaatataa tacggcaact ccagatagca
12241  gaactgtcct cttatctacg cttgtcattg ctgtctgtta gtgacttttc ttaatgttat
12301  aaagtccgta ttctttgttg tgtgtggcca ctgaggtctc tgtctggtta gcttagtggt
12361  taatgactga atacatattt ccttaaatac ctgaaactta caaatcttcc actcttttcc
12421  aagggctata tgtgcatttt ggggcccacc tttaacactc agataagcaa ttaacaactg
12481  ccttaaatcg cttgcacaga ggctcatgat caggtagagg tgagagttta aggccttttc
12541  tagtctttcc tgagcatgcg cacagccctg aacatgtgtg tggccttctg gttccaagga
12601  tacacggagg ttttcaaagg ccttatggag aattcactcc tccaattttc cttttaagtt
12661  tattgatcgg tctactattt cctagtgctt ttagggcatt tccctagaga agttcagtaa
12721  gcattaaaat tttaaggtta ctgatgcata ttgctaaaca taatgctgta tttaactgca
12781  ctcacattag ggtgagctac ttccctggac tagctgagac ttccagtttt agaatcaaga
12841  ccaaagggtt tccagacatg gggcattcag ttctaaaacc atgaatgttc tgggtgaact
12901  aggttgagtt ggttggtctc acacacttta gcactgtatc aatgacctat ctaaagtggt
12961  tccaaaggga caatttaaaa acatctatta ctgaactcta ggctaaacga aaatttattt
13021  ggacacaaag tgtgaagtga ggctctaagc taattttatt tccctaaata actaggtaat
13081  atgttcaaca atattaatct tttaagtcca ttccctacca ctaagttttc cttatcatac
13141  tgtaacttaa ataaattaca gtatatttta caatgaaatg tagaaatttg aacagctcag
13201  ctgaaaggat tttgctaacc Rtaaaaatca tgaaaccact accccagaca agacatgtat
13261  cttttccacc gatgttcaac tgtgcacttc tatagtcagt cccaatcact ggcaaccact
13321  ttctgactta cagcaccaca agttagtgtc acctgctctt gattatcatg taactggaat
13381  tatatagatt gtatttttgt tatctgtttt gctcaacata cttgttttgt gtttcagtag
13441  ttgttccttt taattgctga gtagtattat attgtgtaaa cataccacaa tttagcaatt
13501  gtcctgctga tggaaacctg ggtcatttca aggtctagat tattatgaat aaggttgtta
13561  agaatgttct tgtcaaagtg tttttgttga tgtatgcttt catttgtcct gggtaaaacc
13621  tgggtgataa gattagatgt atgtttaact tcttaaaaac cttcccaaaa gttcttgtag
13681  tggtatacca tttacacgtc cactagtagt gaatgagggt tgtagctgct ccacatcctc
13741  atcaatgaat gttgttattt ttttcttttg ttgttcaact aagtgggata tagtatttca
13801  ttacgacttc aagaatacct atttttatct aggcataatc aatcttcttt agtgtatagt
13861  taagttttgg caaatgtgaa tttcgtgcag ccacccacaa tcaaaatata gaacaattac
13921  atcactctta ttcaactccc tacaaacctc ttcattcact gcggtttgtg cacttctctg
13981  atgcctaatg ttaagcacct tttcatgtgc ttactggcca ttcacatatc ttttgtgctg
14041  tacctgctca agtctttggt gcatttttaa ttgggttatt catcttttta ttgttgattt
14101  gtaggagata tttatatatt gtggatacaa gtagtatgtc atgtatgtgc tgagaatatt
14161  ttttaccact gtgtggcttg cctatttggt ttttttcttt ctttttttt gagacagggt
14221  gttgctctgt tgcccaggct ggagtgcagt ggtgggatca aagctcactc actgtagctt
14281  caaacaccta agctcaagtg atcctcctgc cttagcctcc caagttgctg ggactacaag
14341  tgcccactac tatgtctggc taattaaaaa aaaatttttt ttttaaagat ggggtcttgc
14401  tatattgatc aggctggtct caaacgcctg gcctcaggcg atccacctgc ctcagcctcc
14461  tgagttactg ggattataag caagagtcac gtcacccagc gtatttggtt tcttaatgtc
14521  atcttcttta tggttagtgt tttatgcatt ttgtccaagg aactgttacc tactctaggc
14581  ctgtgaacat agtctctgat attttcttct aggaagatta tggttctgag ttttatattt
14641  agatctataa tccatctaga attaaatttt atgtctggag taagactggt ttaggtttat
14701  aattttatat acagatatct aggtattata gcaccatttg ttaaaaagat tttcccttct
14761  ctattgaact gacttggtgg ccctctggta ctatttaaga aatttgtcta atccatattt
14821  actaagattt tcttgtatgt tttcttttag aagtttagtt ctatggtcca tttcaagtta
14881  atttttctat acaacatgaa gtaagaaatg agattcttct tcttttcctt gtgtctactt
14941  gttctatcaa ttaaagaaga gggcatttaa atctccaaat gtgactgtag atttgtccat
15001  ttccttcctt agttctgata atttatactt catgtatttt gaaagtatta ctaggtatgt
15061  cttcttgatg aaatgacact tttatctttc tcgctgcttt tataattatg aaatctatgt
15121  ataataacat tatttgtctt aaattctatt ttgtctcata ttaagacagt agctttagta
15181  tgctatttgc atagtgtacc ttttttcact tttacttaca acccatgaac ctttatattt
```

FIGURE 1-E

15241 aaagtaaatt tctgataaac agtatatagt ttggtcttgc ttttttgaac caagccgaca
15301 atttctgtat cttaaaatcg tagcttttaa gattagataa agaaaattta agctgctata
15361 gtttcagatt agatatagat aagatatagt taagatacag ttataggtaa ggtatagata
15421 agatgtagtt tcagattaga tatagataag aaaattgaag ctgctacagt ttcagcttct
15481 tgggaggctc aggcaggaat atcgtttgag cccgtaagtt tgaggctata atagtgctct
15541 atgattgtgc ctgtgaatag ccactgtact ccagcctagg caatgcactg aagccctgtc
15601 tcttaaaaaa taaaagaaag aaaaaagaaa agaaaattta taaacttgaa ttacaaggag
15661 acagagctgt aaatatgaaa ttggggttaa cagacatgaa aaacagattc acatccactc
15721 ataggagttt agggacagag tacatacaac atgggaaagg caaaatttga aaagacaatg
15781 gctaagtctt cagaattgat gagagaagtt tttagaattg atgaaagaca tgactcctct
15841 gactcagtag gcataatgaa tcccaaacac aataaataat atgctagaca tcgacaaagt
15901 ttcacatttt agtgaaactg cagcacatca aacacaaagc atcttacaag cagccagcaa
15961 acccccacaa gtcacctacg aacaagacag agtagacttc tttcgttaaa tagaaacaac
16021 acgccagaca acataaaaca gtatcttcaa aatactgaga ggtaacagct atcaacctgt
16081 gtaattctag attgaactaa cttttaaagt gaaataaaat aaagatactt tcagataaag
16141 tctagaatat ttacgattaa ggaaccattg ctgaaagcac aaatcaaggg cttgcacact
16201 aaggtgcatt tagatatctg aagggcattt ttgcttttta caatgactag ggagttttac
16261 tagcatgtag catgtatatg aaagctttta aaaaagggaa attttcggcc aggtgcagtg
16321 gctcacgcct gtagtcccgg cactttggga ggccgaggtg ggaggatcac gaggtcagga
16381 gatcgagacc atcctggtta acaaggtgaa accccgtctc taaaaatаca aaaaaattag
16441 ccaggcgtgg tggcgggtgc ctgtagtccc agctactagg gaagctgagg caggagaatg
16501 gcatgaaccc gggaggcgga gcttgcagtg agccaagatc gcgccactgc actccatcca
16561 gcctgggcaa cagagcgaga ctccatctca aggaaaaaaa aaaaaaaaaa aaaaaaggga
16621 aattttctat gaagagcaag acagtcctat acaacaaata actgtacaat ccaaatgctg
16681 ttagtggtcc actgagaaac acagaagagg gaagatacaa gatgggacgg taaacaaaaa
16741 ttgtatgacc tgtggattaa cttttaataa acaaaaataa aaattacaat ttatgggtat
16801 aaacatgaaa cagaagtaaa atactagaaa tggcatgatt agaattaaag tttcccaaag
16861 tctgtgtgtc acagtagaac tataggtaag ttttagacaa aaattgtgta tgggtaacaa
16921 cataagcaca caaattgaaa gtaagaagag gagaaaagat gacacaaata tcaagccaaa
16981 aaaaaaaaat cagccacatt aatagcagac aatatagatt ctagggtaaa accactatta
17041 taaatcaaga tgttcattac ataataacaa ctataaaact atgtgtaccc aataatacaa
17101 tttcaaaaga catactaaaa aaacctaatg ctatgaggac caactgttga acccatgttt
17161 gggtagcttt aaaaaaatca tatttggata gctttcaaaa cctcaataga tcaagaaggt
17221 aacagattaa agtataaaga agtacacaag taacaaactt gagttaatag acacatagca
17281 aactttcYag tcaataaata gagaacacat aatttcttca agcacattta taaaaactga
17341 cccctttcta taaaagaaca cctattatat aatatgttct ctgacactac aattctacaa
17401 ttaagctgta aaatgttaag aagataacca agatatcccc attactttag gatataaaaa
17461 aacagaagac cttgctttaa aaataggtcc agtgaaaata acaaaaacgt gatcttgact
17521 aacgtgtctt accgatataa tacgatctcc ttttctgagc tctccactta gatcagcagg
17581 tcctccggct aagataaagg aaataaatat tccttctcca tcttctcctc ctacaatgtt
17641 gaaaccaagg cccgttgagc cacgatgaag aacaactttt ctaggttccc taaaaattaa
17701 aaaaaattga gtatcttgga aattttatat aaaacctgac attctatcag tgagaactac
17761 ctactgataa tttttatgtt atcaaactac aaagaaagct agacaaaata taaatggttt
17821 ttctaaaact acgagataaa gagattctaa gcctacaaat caaaggaaaa aatctaattt
17881 tatataaaag caaaaaacct gtgttatata aaacagaaat gaaaccaaaa cagaaaaaaa
17941 acgtggaaat ttaaaagtat tataagagct tatcaggaaa tacacatact catgaaaagc
18001 tgatgttgca cactttattg ccttcaaagc aatcactgcc cctacttaga agtttaaaag
18061 gcaggtatcc ttttgtctta gagaaggcag gccctttctc cagtctcaaa gggatgaacc
18121 ataaccaacc taaacaagtt attgcaatcc catgcccctt tgctgaggtg agagtataac
18181 cccattctaa cctctggtaa atgaagggaa attttattgt cattttctag agacagacaa
18241 accttatcaa ggacaaagca tctttgctac cttcctttcc tcaaaaaaac gaaacaaaat
18301 aaacaaaacc ccacacgact gaataataca atccctttac tggtttggac atattataat
18361 caggaaggtg aaaccaagag tataccgaag acaccacagg cctaaggttg ctataccaac
18421 cttggaaatg cccatcttca ggctttttat atgaaataat taaaacctat ctgcaggggc
18481 acaggtaggt tttatcatat tagtgatgtt ttatttccta aattagatgg tgaattccgg
18541 atttactgtt cccattatgc tttaaaatat gtatgaatta gaattttata ttaggaaatt
18601 atacttgatt aaaaaatact cagctgaaat tcaacagtat ttccagggga aaatacactc
18661 tgattttcaa ataatgtaag cacatgaaaa ggtgctcaac atcatcagtc actagggaaa
18721 gagatataaa aataatgaga tacttcaggt ctactaggat ggctattaaa aaaaccacaa
18781 aacaaaaagt aacaagtgtt gatgaggatg taaagaaatt gcaactcaca cattactgat
18841 gagaatataa aacggtgcag gcactacgaa aaccagtttg gtggttcctc agaaagctaa
18901 catagaatta ccatatgact cagcaattcc acccttaggt atatatccca agtaactgaa
18961 agcagtgatt tggacagata cttgcatgcc agtgtttatt atagaattac tcacaataac
19021 caaaaggtga aaataactca agtgcccatc attagatgaa taaacaaaat gtggcatata

FIGURE 1-F

19081 tgtataattg aacagtattg tcataaaaac aaatgaaatt ctgatacatg ctaacaaatg
19141 agtgtatctt gacaacataa gtgaaatagg ccagtcacaa aaggacaaat attatacagt
19201 tccactttta taaactatcc agaataggca aattcataga gacaaaaagt agattaaagg
19261 ttaccagggg ctgggaagac agtggaaggg agaattactg cttaatggtc acagagtttg
19321 tctgaagtaa tgaaaaggtt ttggaaatag tgaagggttc cgaaaattgt gaatgcaatt
19381 aacactactg aattgtacac ttaaaaatag ttaaaatgtc aaattttgtt atatatattt
19441 cgctacaatt tttaaaaact gatgtaatat accacaaatt gtatacttta aacaggtgaa
19501 atttatgata tgtgaatcat atctcaataa agctgttaaa aaataaactt tagaaccaaa
19561 atgtaggtat gttgtgattt tttttttac ttttttgata ttggtaacat ctgaaagact
19621 gcttaaagtc aaattgtgaa gaacttataa tgttggaaag attttatact tcattattac
19681 aaagtagtgt gattatcaaa agggagtggt tcatacttaa aagtccaatg caatattcta
19741 gacaagagac tcaagtgaag aagcatgagg aacagtaatc aaggtgcaaa tataacttat
19801 tttttagttt gtaaaatatg caaagagatt aaagactaga taagccattc actattacag
19861 tttccctctt tacggcctta aataggcact attagaaagt aataaaaata aatggcaatg
19921 aaaggtcact ctagaagcac tgcctgaaga ctagcagcct tggatattcc catcacaaac
19981 aaataagaac actattcttt ctgctaattt tcatcccaaa cacaattact gacaacctat
20041 taagtttcca acattgctaa ttctttatga aagaaaagag acaaacactc ctatctgtcc
20101 taaagatcac tgcctagaat caggagtctg ataagtaaaa aataataata atgctaacaa
20161 taatgtaatc aataaacgtt aagaacagac attacttagt acacatttta atgttgaatc
20221 tgtaatataa aagacgtgta caaattacaa agcaaactct atagtgtctt attactatat
20281 tgaataatat gaaaaagatc tacaatgctt tttcaccaat ttttttctac ctcattagaa
20341 ttctttggga ttaaaaagac agttaatcta tacatttcag tgcaggtagt aattttagat
20401 gaaagtaaat tttgtgtgtc agaatatcaa ggatatatag caaacaaaca aacccaccca
20461 ccagccaggc aaaggggatc acgcctgtaa tcttagtacc tagggaggca gatacgggag
20521 gatagcttga gtctaggagt tggaggctgc agtgatgtac gttcatagca ctgcactcca
20581 gcctgggcaa cagagtaaga ccctcctgct aaaaccaaac caaaccacca cttattgaat
20641 tctgaacaca aatcaaataa ctgccatatt tttatggtat atattagata aggaacatat
20701 aaaatgttac tttaaaatat gcataagaat tttcttaact tagttttact aagctaattc
20761 ctaaggacaa tttaccaagc ctcaaagaaa agcagtatta attttaaaaa aggagtggta
20821 atttatttgt aaaaataaaa catgtatatt tcaggctctt caatgaatcc tcctatggaa
20881 aaaaattaac ctttaagctc actaactgtc aataaaattt tttagtccta aaaattgtgg
20941 ctatcttaca tggctgatta aaattcaatt taatagttga ttttatgtaa gaaggataaa
21001 tgttaacttc cttaccttgt aatttcatca tctccaagta ctgctttaga aactggggag
21061 tatctggctg gagatgctgg tgtctggccc aagaaggaag atgggctaac atggttatca
21121 acaggctgag aagaagcttc aaaataaaca aagtgaaaaa tacttcaaac acgaaacaag
21181 ccaatcagta ttccatttat gagtgattaa tgtgtaattt atatgcactc ctttatatat
21241 cagaatttgg tagagaagat ttactcatca gccaaaaaac tggacattat gttgcccagg
21301 ttagtctcaa actcctgtcc tgaagtgata ctcccacctc agcctcccaa agtgttggga
21361 ttacaggcat aagccaccac acccagatta cttaaagaat tatatacagt ccaaatttga
21421 ttgtgaataa taaaaatcag aggctttcca ataagtgcga aaagactatg tccttatact
21481 agctcagatc tgtcaaccta atttacatct atgcttttaa attcacccat agaagaataa
21541 aaacctggta aaaagcaaaa acgaaaaaca agcaaaaact gtcgtccagg cgtagtggct
21601 cacgcctgta accctagcac tttgggaggc caaggcaggc ggatcacctg aggtcaggag
21661 ttcgcgacca gcctggccaa catggtgaaa tcccatctct accaaaaaaa tacaaaaatt
21721 agccagactt ggtggtgtgg cctgtagtcc cacttactcg ggaggctgag gtgggagaat
21781 tgcttgaacc tgggaggtgg aggttgcagt gagccgagat cgcatcactg cactccagcc
21841 tacgtaggtg acagagtgag atgccctgtc tcgagaaaga aaaaaaaaaa aaaaaaagca
21901 aaaaccaaac gttggttcac ttcaatagta ataaatacca catataggtt ttccattcta
21961 gcaaaagcta ataacagaaa attatagtga ttcctgacca tgctttctaa agacacaggt
22021 aggtaacaca tggcagctgt agcttacaaa gacataagac acttgaatta ttccaatcat
22081 taccaaaaca cagaggaagc aatatttaac tttcttgagg cttcaactat gataaagtta
22141 caaagcactt caaaagtagc tgtattattt aattatcaag cattaatctc ttttttatta
22201 aattagagca tatcttctat ggagggaagc agcatactac gcactggagt acaaaaatgc
22261 aggaattatt agttcaaatt actatagtgg ccagataggt aacataaagg aataaagtga
22321 actggatgaa agacaacagg aaatgactga aacgatagta ttttagagat gcagtgtatc
22381 tattgatatt acaggtttgc agtatccaac agcaattgtt tcctatccag ttcatatata
22441 agatgctcgt ttgtatttga gccaaaggac tttctaccaa tggctcttaa ctttgaaagt
22501 ccaaagtctt tcctgggtgg tcagtaagaa tatgggattt tcaagtgaat tggtatgtag
22561 ccctcaatca atagataccc acgtgaacct ctacaatcac tagcctgtta aaaatccaga
22621 gtttactgat ttttgactct taaattcttt tgctgcattt tcatatttag atgaaacaaa
22681 aaaaacaact agacaagaaa tccagtcaaa tgcccaaacc agaaaatata cattttccct
22741 gacacatcca gactatccct ttagtcaatg catccttttct ggggcagtta atctcacatg
22801 taccacatac tctcagacaa cagagactaa aaattaatgt tcctatgaaa gaaatgacgg
22861 cctcaaagaa ggatcagata aaacagtact atttcttata ccccaaatct tatgtaaaaa

FIGURE 1-G

```
22921  ggtcccccag agaagcacaa gagtgcctaa ttcatttttc ttatatatta agatgaatga
22981  ttcaacacta taagatttta tagaaaagca agtagaagag cttgaagtga gaatgggaaa
23041  agactggtgt ccaaatcaga agtagcaagg ccatgcagct aaagagaata taagttaaaa
23101  gcgagaaaat atatactgta aaaaacagga acattttaat gtcggcagag aaactgtatt
23161  ctctgttaga ttcaacagac ttctttttt tcctactttt attcctgagt atcatatttt
23221  gactacctta ttttggttat aaacattgtg gctctttatc ttctatagtt catgtataac
23281  ttctatgatc ttccctattt cttttactat tcttgctaaa aatatttttc taactagtac
23341  aaaaattgtt ttccaatttt gacatctgct cacttgcaag gctgcctctg tgagaaacat
23401  cccctccatc aatgcttgac aatcaccaca aaaaaaaaaa aaaaaaaaaa tcaaatgtta
23461  tctagtcctc ttaggtaaca agagaaagaa aatctgaaca ttcttgtgtc tgctaaggaa
23521  aatgcattca gaactaaaat ccttcccaag aaaactccag ggccagacag atagatagct
23581  tcattggaaa attcaacaaa acatctagag atgaaataac attctgaaac attacaagac
23641  agaaaaattt caggacacta tcgttaagtt catgcttgtg aacacagatg taaacctaaa
23701  caaaatataa gcagttcaaa atcaatgata tataatatat attataagga tgtgggtttt
23761  ttcccccccac aaatgaatgg tttagcgtta gaaaatcaat gtaatttatc acgttaacaa
23821  aataaaagag aaaaatcatc tccataaata attggataaa attcaatact tacaggataa
23881  aaactcccca taaaccagga ataaagatta actttcttaa tacaaaaagg gtatctacag
23941  aagacctgct ggtaacacca tatacaatgg tgaaatactg tgtcccccag tttgtgaaca
24001  tgacagtgat atttattatc accatttctg ttcaacactg aactggaggt cttaccaatg
24061  catcagacaa gaaaaaagaa aatgaacagt ataaagatag gaaagaagga agtaaaattc
24121  attgttcaca gatgacatgg ttatatacgt agacgatctg aaagacctat aaaacctaac
24181  attaacaaat ttagcaagat tgttgactaa gcaattaata tacaaaacag atatttccta
24241  tatatcagtt aaaattagaa acatttacaa agtagtacat ttataatagc atgaaaaaca
24301  tcaaatagcc aggactaaat ctaacatata aaacctctac tatacaacac tgcagagaga
24361  atgtaagaag gacttcaata aaagaagaga tattcatatt aatggactga ttaagaaact
24421  caatttaatt cttcctaaac cgatctgtaa tgtttttgat cctaaaatta aaaaggaaat
24481  tcaaagaccc aacataatct tgaacagtaa caaaattaag atttacttta catcaatatt
24541  tattataaaa ttatgataat taagagactg caggtggtac aaagacaaat agttcaatga
24601  aacagaagag cctagaaata ggttcatata tatgtggtca cttaaaagaa aagcaccaat
24661  gcaattcagt ggcagaaatg gtctttcaat aaatgatgct agatcaatta tatatctgca
24721  tattaaaaaa tctctcatta catgcaaaaa ttagatcaaa atggatcgga cctaaatgtg
24781  aaaggcaaaa taataaagct cctacaagtc ccttatgacc taaggatagg aaaaatttca
24841  acaggtcact aaaagcacta ccttaataga aaagactgat aaggaattaa aaaatttttat
24901  ccatcaaaaa gtaccattgt tttgagggaa aacataaact cagtgaagac atctgcaaca
24961  gatgtaactg attagagtta tttctcaaat acataaaact tcctttaaat caataataaa
25021  accaatggaa aaatgtaaaa agatttgaac agacattgca ttgaacaaga gtcaaagcaa
25081  taagcacagg caaaaatgct gaaaataatt aatcatcagg aacaaccagc aataaatgaa
25141  taaataataa tcattaaaat catgagataa ctttatacat atacatacta ttcataatta
25201  aaaagatgga taataccaag ggttggtgag gatgtagaaa aactggggtc ttaattgcct
25261  cctatactgt tgaaattgtt cagcagtatc cgctggagac taaacatatg cctaccctgt
25321  aacacagcaa cctcactcca tgagaaatga ctgcttatgt ctgacaagga tgtgcaaaaa
25381  catccacagc agctacactc ataaaagatc agaactgtaa gtgactcaac agtaaaatga
25441  aaaaaattgt tgtacactta tacaatggaa taatacatag cagttttaaa aagccatgtg
25501  actgatcatt aagctaaatt cttaaaattg gtacccttac tgaatatatg ttttaattca
25561  attaaaaatt aaataaagag gtcaggtgca gtggctccta tctgcaatcc caacactttg
25621  ggggcactga ggcaggatca cttgaggtta ggagcttagg caacatagtg agaccccatc
25681  tctacaaaac aattttaaaa attagctggg catgctatcc tagacacttg gRaggctgat
25741  gtgggaggat ggcctgagcc caggaattca aggctacaag gaactatcat cgtgtcactc
25801  cactgcagcc tgagcagctg agtgagatcc tgtctcaggt aaaagaatct ttttatagac
25861  tttccccatt tctttactgg gtatgttttt ttatttttct tacagagtct tgctttgtcg
25921  cccaggctgg actacagtgg cgcgatctca gctcactgca aactccgcct cctgggttca
25981  agcaattctc ctgccttagc ctcccgagta gtggggactg caggcacttg ccaccatgcc
26041  cgggtaattt tttgtagttt tagtagagat ggggtttcac tgggttagcc aggatggtct
26101  tgatctgctg acctcatgct ccgcccgcct cggcctcaca aagtgctggg attacaggcg
26161  tgagccaccg tgcctggtct ctttttttaag agatgggtct cattatgttg cccaggctgg
26221  agtgcagtgg ccagacacag gtgcaataat agtgcactat aacccccaaa ctcctgggtt
26281  caaacaaccc tcctgcctca gcctgccaaa taattgggac cacaggcacg caccattgtg
26341  cctggctttt ttccttttttg atgtgaagaa gtcttaacat ggtaggaaaa tcagctctcg
26401  gtgatattaa gcagagaaaa ggagcagtgt tagagaggta tcttaagact gtaggacgat
26461  ccctcctatt tccttcagat aaggagtatg taaacatgtt cgtgtgctga tggtaatgat
26521  ccagtagaca gaggatttga ttatgcaaga gagagaaggg acagctgcaa gaacaagatt
26581  ctctgcagac aacaaacaat gacattcagt atacaaatgg aagacctggt tttcaacagg
26641  aacaatgaca gtgactccca ttaaaacaca agtgaaagtc gagtttcggg gtaaaaatat
26701  aattacatag gtatatctga aagtggaaaa ataaggaaat tctcttctgg gtgcttatat
```

FIGURE 1-H

```
26761  gaaatatgaa gtgagatcat cattcatgaa tagtagatgg caagtgaagg tttgagggga
26821  gagaaaaatt ctttagagtg ggagagtgaa ttagggaaat gtagtaggac tgcaggcaag
26881  agtaaggcac tatctatgat ttatggccat acgtaaaaag caSaattttg tgcttttctc
26941  catggttctg ctttttagga accataaatg aaatagcaga gtttgcctgc accagtgttg
27001  ggggattctg ctgcgaagaa gaaggggtca aaggacttga agatgtgcaa tgaagtgaac
27061  atgagaaacc atggaatcta agctaataca aatagaaaaa ttgagacaag gggcaacaaa
27121  ataataatgt caaatgactg aaggtcaaaa tgagatggaa ttgctagaat agaggtgaat
27181  gaactgaggg ttgtagtcac acaatgggat gatcaaactg gtattttaga ggtgatgaaa
27241  gttaataatt atagataaca aaatctaaag tattatctta gaagcaagta actgagttga
27301  agtggagggc aggatagtca gaaagagatg aagaaaccaa aaagtcaggg tgttggatga
27361  atcattagtg tggtttttgg acatgaccag gaatgacagc atgagtaaca gtggcaaaac
27421  gacaatgggc cttcgcacta aagtcttcag tgactacagc cagactgatg aagagaacac
27481  cagatcacca cagccaggga tagagagtaa gctgaatagc ctgacagcag ggtgactgtg
27541  caggcttcag aagaactgac aagtaaaact cacttatgta aaaaaaattt aaacacaagt
27601  gcttttcaaa aaacacaaac agattgttta tgaattgcat atgcagaagt gtacaacgaa
27661  acccacatct acagttgcct aagaagggaa gggactggaa gcaaaatatt atgatcaagt
27721  tgaaactgca aggtgaatgt cagcttttc ataatgcttt attagttcaa taacagatgg
27781  gggaaaaagt aacataatca gctgggctcg gtggctcatg cctataatcc cagcgctttg
27841  ggaggccgaa gtgggtggat catctgaggt cgggagttcg agaccagcct gaccaaacat
27901  ggagaaaccc cgtctctact aaaaatacaa aattagcagg gcgtggtggc gcatgcctgt
27961  aatcccagct actgaggcaa gagaatggct tgaatctggg aggcagaggt tgtggtgatc
28021  cgagatcacg ccactgcact ccagcctggg caacaagagc gaaactccat cttaaaaaaa
28081  aaaaacaaca taatcataat cagggcacta atactcaatt cgtggaacaa ctgtcacaat
28141  gtgcacatgg ttattagata ggcagcattt aaaataagat acttgaattg atgaataaaa
28201  tggctcatta tttaaaaaat acacaaagcc tttatataaa gtttatgtgc tagaggaagt
28261  atatgtaaga atttcaaata agtagcaagg ttcttttctt tgacacaaaa gaagtataag
28321  acaccatccc tgtacatcag acaggttaaa tactattaag gaaataattc aattatactt
28381  gagcaattaa taaatcaatg agcagataat gaagatacat tactggaggg cagtatgtag
28441  atttcaaaat gtcatgtttt taactgataa tgattagtaa tataatgaat cttgacagtt
28501  ctaaaattgg aagcactgtt actttaaaaa tcaccaatat ttctaaaatt ctacaattta
28561  aaaaaggagc actcaaaagc aggttatacc cagtacgttt aagatcttta ttatttacgg
28621  gttcttcaaa cattaactca atgcaaagaa caaatacaga tttcattttt ccaccaatac
28681  aaacacatta aaaaatatac ttaaatctct tctcagctta tatatttaaa aaactgaata
28741  taaaatggcc aggtgcggtg gctcacgcct gtaatcccag cactttgaga ggccgaggca
28801  ggtggatcac ctgaggccag gagtttgaga ccagcctggc caacatggcg aaaccccgtc
28861  tctactaaaa atacaaaagt tagccagtca tggtgtcgca agcctgtaat ccctgctact
28921  agggggggctg aggcaggagg atcgcttgaa cctgggaggc ggaggttgca gtgagctgag
28981  attgcaccac tgcactccag cctgggcaac agagcgagat tctgtctcca aaaaaaaaaa
29041  aaaaaaaaaa aaaaaaaaaa aaaatatata tatatatata tatatatata tatatatata
29101  taaagtaaga aacctaaaaa tacgtaagta ctttaagaac aatttaatca actgcaacac
29161  aactgaactg catacaaata taagcactag aacctgaaag tacaaagata aatagtatct
29221  ctcctcatgt acctaagagc aaagaaaatc cctttaattt tagatatatt gtaaatcagt
29281  gtttctcaaa gtgcaggcca tgaactagca gcaccagaaa cctcagaatt acagacatac
29341  ctcagagata ttacagttcc ataccacagc aataaagcga atattgcaat aaagcctgtc
29401  attataaatt ttttggtttc ccagtgcatg tacaagttat gtttacacta tattaagtaa
29461  ggaatagcat tatgtctaaa aatacaatgt acaggcctta atttaaaaat actttattgc
29521  taaaaaaaaa ccgctagcaa tcatctcagc cttcagcaag ttataatgtt tttgctggtg
29581  gagcatcttg ccttaataat gatagccttg atactgaggg tggtagttgc tgaagggtgg
29641  ggtgcctgtg ttaatttctt aaaataagac aaaaatgcag ttggccatat ccactgactc
29701  gttcttttac aaaagatttc cctgaagcat gtgatgctgg tttgttagca ttttacccac
29761  agtagaactt ctttcaaaac tgaagtgaat tctcttaaac cctactgctg ctttatcaac
29821  taagtttatg taatatccca aaatcctttg ttgccgtttc gataatgttc acatcatctt
29881  cacctaccag gattagattc catctcaaga aaccactttc tttgattatc tataagaagc
29941  aactccttag ttgttaaagt tttatcatga ggttgcagca attcagtcac atcttcaggc
30001  tccacttcta attctagttc tcttattatt tctaccaatt tgcagtaact tctgccacta
30061  aagttttgac ctgttccaag tcatccaaga gggctggaat caacatcttt gaaactcctg
30121  ttaatgctga tattgtgacc tcctcccatg aatcatgaat gtccttaaag gcatctagac
30181  tggtgaatcc tttgcagaag gttttcaatt tactttgccc agatccatca gcggaatcac
30241  tacctatgac agctatagct ttataaaatg tatttcttaa atagtaagac ttgaaagtca
30301  aaattattcc ttgatccatg gactacagag tggatgacaa gttagtaagc atcaaaacaa
30361  catcagtctc cctgcacact gccatcatag ctcttgggca gctaggtgca ttgtctcaga
30421  gcactaatat tttgaaagga gtgtttttt tttgtttttt tctgagcagc aggtctcaat
30481  agttggctta aaatattcag taaaccatgc tgtcaacaga tatgctgtca ttcaggcttt
30541  gttgcttcag ttatagagca caggtttcat tacagttata cagaacaggc agacaggctt
```

FIGURE 1-I

30601 agcataattc cttggatttt ctggactggt aaatgggcat tggcttcaac tgaaaatcac
30661 cagctgcatt agcaactaac aagaaagtca gcttgtactt tgaagctttg aagtcaggca
30721 ctgacttatc ctctctagct gtgaaagtcc taaaaggtgt ctttttccag tagaaggctg
30781 tttcatctac attgaaaaat tgttgtttgg tgtagccact tccattaagt atcatagcca
30841 gatcttctgg ataacttgct gcagctccta cagcagcact tgctgcttta ccttgcactt
30901 tcatgttata gagatgactt ctttccctca acctcatgaa ccaacccctg ctagcctcac
30961 atttctcttc tacagcttcc tcacctctct cagccttttc agaattgaag cacagttagg
31021 gtcttgttct ggattaggct ttggtttgaa ggaatcttat ggttggtttg atctatctag
31081 accacttcaa actttgtcca caacagcaat cagattgtct tgctttctta tcattagtgt
31141 gttcactgta acagtacttt taatttctt caagaacttt tccttggcat tcacaacttg
31201 gctaactgca gcaagaggtc tagctttctc cctttattgc cattgaacat gccttcctca
31261 ctaagttaaa tctttttgac atataatgag aaatatgcaa cttttcactt gagcactcag
31321 aggtcattgt agatttatta actggcctaa tatcaatatt gttgtatctt ggggaatagg
31381 gaggcccgag gagagcgagg gagatagggg aatgactagt catcagagga gcagtcagaa
31441 cacacacatt tatctattaa gcttgctgtc ttatatgagt gtggttgtgg cactccaaaa
31501 caattaaaat agtaacatca aagactagtg atcacatatt actgtaacag atataataat
31561 aatgaaaaaa gtttgaaata ttgcgagatt taccaaaatg tgacacagag acattaagag
31621 acatgaagtg agaacatgtt gctgaagaaa aaaggtgctg ataaatttgc tcaattcagg
31681 attgccataa accttcaatt tgtaataaat gtagtctctg tgaagtgcaa taaagcaaag
31741 cccaataaaa tgaggtatgc ctgtactttg cttgttaata gaatgaacgg cttgttcaag
31801 caaatttctg tgctcaagtg gaaaacttag aacaaaacaa tttcaatatt tgcaaatgtt
31861 ctttagtgac tgttacttac ataaaagttt gaaaatcctt taaatgtaac tactactata
31921 aaataataaa ggtgaaaagg atccccccct tctaattata aaaattttga cttaaagtag
31981 attttaaaaa atgagtagat taacatgctt aattgtttct ttaaaaatat ttgcataatg
32041 tttaaacttt attatactga gaacatttca ctaatggcat cacaactgaa gaagtaagat
32101 aaatttaagc aaacgtatgc taacagactt acagttggtg atatcaggtg gtgcatagcc
32161 atcattcata tacatacttg tgggttttgc cactttcaaa taaacaaaat cagatgtgtt
32221 ctttaaggca gttactgctt cttcatgagt aacttcttct aaacatacgt tattcaccta
32281 aaaaaagtcc caaaagacat ttatgacata ttcacttgtt ctctgagttt tggaaacaat
32341 ttcacgaaaa gaaagggaaa ataagagagt ggtttaagga aaataaaggt atgccgaaag
32401 aaaataccca tgtttgatgt ctatgatctc agtaagttgt ttatcatgac atgtgaagtc
32461 gttgaatgtt aacagaatag cccatcttga ttccagtctc catacctcct attctcccag
32521 tcttgatttt ctctcattct taggcattta ggaatgtgtg tggatgcgtt ttagacattt
32581 ttatctttaa agaaatgaaa acatactaaa atttttgttt ttatgttagg atttttaatg
32641 taaggtacct tagagactgt tccatgttgg gacacaaatt aatagaatta ctacattgta
32701 acaaaacagc tgcacagtag cccattgtgt ggctatatca taatttatta aagatcatat
32761 ctgttaatat tcccatattt agattggaaa actattagaa cacaattgag cacaaatatt
32821 taaattgttc tcctaaatgg ttttagtaaa tttatactca ctaccacaga atatgagagt
32881 gcccatctac ttagaccttt gccaacatta aacattatca actaatttaa aaaatctgta
32941 aaatgatgca tcttttaatt tatacatcaa ttagttttaa atttctactt tcgtaatact
33001 caagttgatt ttttttcaat tatttttgg ccattcacct atcttctgta tagaaatata
33061 actgattttt aggtagtgat gtatcctgca accaatacta aactaattta ttagttctat
33121 tagcaggttc ctaggatagt ctatatacaa tatgatgaca tcagcaaata tagttttatt
33181 tcttcctttc ttttttttt ttttttttt ttttttgag acggagtctc gctctgtcgc
33241 ccaggctgga gtgcagtggc gggatctcag ctcactgcaa gctccgcctc ccgggttcac
33301 gccattctcc tgcctcagcc tcccaagtag ctgggactac aggcgcccgc cactacgccc
33361 ggctaatttt ttgtattttt agtagagacg gggtttcacc gttttagccg ggatggtctc
33421 gatctcctga cctcgtgatc cgcccgcctc ggcctcccaa agtgctggga ttacaggcgt
33481 gagccaccgc gcccggccta tttcttcctt tctaatctga ttttcttttt ctttttcttt
33541 tcttattgcc ctggatagaa tctcaactat aacactgatt agaagtggta gtttatgtct
33601 ttctagcaat ctgtcaattt catctaagtt acctcaggta ttagcacata gttactcata
33661 atagtctctt gttttccctc tttcatgctt gctttcagaa atttcagcct tctctctttt
33721 ttcttggctg gtctgaaggt ttgtcatttt gttgattttt ttttcaaagg accaactttc
33781 agttttgtct gtcttcttac atcttttcta ttacctattt cattaatttc tgctctaatc
33841 tttattattt ccacccttct gtttgcttta ggtttggatt gttcttctct ttgatacatt
33901 ttttttttt gcaaataagt aaaacattat caatgtttcc gttttaaatt accatgtatt
33961 aacaatatta actttaatat atttttcact ttttcctaat ttgtaccttt atagttgact
34021 attgcagttt ttaatatttt cacaaaaaaa tgtaaaggtc acagaataat aatttccctt
34081 cactgatgat taacattgct ttgcagagtc tcagcagcat ggtcattttt atggctccac
34141 attaccatgc aaagcaagga ctgcctgcat tttgaaatac atattaaata gctgtggcat
34201 actgctgaaa tacacattaa atagctgtgg catactgctg aaatacacat taaatagctg
34261 tggcatactg ctgaaataca cattaaatag ctgtggcata ctgctgaaat acacattaaa
34321 tagctgtggc atactgctga aatacacatt aaatagctgt ggcatactgc tgaaatacac
34381 attaaatagc tgtggcatac tgctgaaata cacattaaat agctgtggca tactgctgaa

FIGURE 1-J

```
34441  atacacatta aatagctgtg gtatactgct gaaatacaca ttaaatagct gtggtatact
34501  gctgaaatac acattaaata gctgtggtat actgctgaaa tacacattaa atagctgtgg
34561  tatactgctg aaatacacat taaatagctg tggtatactt ctgaaataca cattaaatag
34621  ctgtggtata ctgctgaaat acatattaaa tagctgtggt atactgtttc tttggtggcc
34681  tacaataagc catgcatcct ggtattcaca cccttgtaaa aatctctttc tacactgaat
34741  tttgcttgtc cacgtaacta gctttagcca atgaggtaca gtagttgtga aggaaagaga
34801  agattgataa aaacaatgta ccagtgcttg tcttctttgg aaagcttact tttggaacct
34861  agccaccatc ctgtgtgaaa ctcaccctaa ccacgcaaag agaccactgg ataattaagc
34921  acctggctaa cagtcccagc tgagttccca gctaagagcc aacttgccca ccatatgtgt
34981  cagccaacct aacagtggat tttgtggcct agtacagcaa agacaagttg ttcctgccaa
35041  gccctatcca atctgcagaa ttgtaaacaa atacatgagt ggtgtttttt agccattaat
35101  ctctgtgttt tctcacatat cactagatac ctgacacaag agctgacaca atttgctgag
35161  ggatttaatg tgaaatgtgt aagatgatga tcccaaagtt ttacgccaga gaaaagttat
35221  ttgaccatct accagagtcg tcttctacac attaagaaac agtgaccttg ggagtgaata
35281  gtaaaaagga gaaaatggac aaggagtgag ttctggatta cttcaatgtt tagaggttac
35341  agatatgagg aggcaccaga catcagaaag acagaatgaa gagtggctac taagaagaga
35401  aataccaaga taaagggtga ctattaggca aatgaaaaag tgttagaagg aggaaggtgt
35461  gtcaaactcc agccagataa gtcaattaag atgaggactg agaattgaat gattccagtg
35521  taattaagag aacaaaagcc tcactaaaag ttgatttaag agagagactt gccacaatgg
35581  tatcagcaca ggttgctgtc tcctaccacc aaagccagtt ctagagaaac cataaaggaa
35641  aacatggtag ccaagaaatt aaaaagtaaa actgagaaaa ctagtgcaac agaaggctgt
35701  cttgactgtg tttgtgtgtg agtgatgagt gatggtgaga tggtactggt tatagtttga
35761  ggaaagagtg gtgatcttag ctgtgaaagg ataagttaga gaaacatagt tcaagttctg
35821  ctctttcctt ttcgccatca ttaactaaaa aaagtctcct tgcaacaata aacagtacta
35881  aacaggatgt agaacaaagg aatctttttat atattgctga tgtaaatttt tacaagcagc
35941  ctggaaaaca acaggatggt ttttaaaaag aatgctacta gtaggataat caataatccc
36001  atgacagtca agtaaaattc ttgtacatgt tcaccagaac acacagcagc tcttcccagg
36061  aaaattattc ataataccaa aaaaaaaaaa aaaaaaaaaa gacttaaaat agatttttaa
36121  aaaccagtac ttgctagtat atttttaaaa accaacttcc tactatgcca gatccagttt
36181  atgtgccttt tcagattctg ctaggccttg gctgcttgct cagtggagag tcctgactat
36241  tatactactt tcatcttgtc tcctgccctt cccagttttc tgagcctctt ctggctggga
36301  ctgggcatgg ctgatgtgtt ccacacagtg agggccataa catgtcacaa acctatgttt
36361  gacctgtcat accactgcaa ctctaccatt tttgggcttg ggtgaagtgt tgtgccatgg
36421  agtataggag ctggcttccc tagaaactgc gcccagggga gttgggtaac acaatccaaa
36481  agcttgggac aaataacccc taggggtag acaagagaca ggagaataaa agggccagca
36541  gataaacctt tctttctttt ctctggatgg ctggatgtaa ggtttatgat gcccggatgg
36601  tttttacagt gtctgcctag aagactgtta tacatgacaa gcaacaagct accttataaa
36661  attacagcca gctcagacac attctacctt ctattggctc tcgttcattc tctgcttcac
36721  tcctcatttt tactcactct agccttcttg agatcacaca gccaaataat accttagcac
36781  tttagctttt cttcagattc cgtttcccta aggaacatgg actgagacag ttattgattg
36841  atatggtatg gctctgtgtc cccacccaaa tgtcatcctg aattgtaatc cccatgtgtt
36901  gggggaggga ccttgtagga ggtgattaga tcacggaggt ggctccccat gatgttctct
36961  tgatagtgag tgagtctcat gagatgtgat tgttttataa gcatctggca tttccactgc
37021  tggcacgtct ctctcctgct ggcatgtgaa gaagggcgtg tttgctttcc cccaccctcc
37081  accatgactg taagtttcct gagtctggtg agtcaattaa accgctttcc tttataaatt
37141  acccagtctc gggtaattct tcatagcagt gtgaaaacaa actaatacac tgatgaagac
37201  tggacaaata aattgcagta cagacataca gtggaatagt atgtagagat aaaatgaatg
37261  aactttagct ctaagcaatg atatgggtaa gtctcagtaa agcaatattg agtgaaaaaa
37321  atactggaga tgaaagtctc atacagtgca atagtgtttt gaaaaagctt caaaacaaat
37381  aatactaaac aaaattattt aggcatatgt atataattaa actttttttt tttttgagat
37441  gggctttcac tctgtcaccc agactgcagt gcagtgggac aatcacagct cactgtaacc
37501  tcaacctcct gggctcaggt gatcttctca cctcagcctc ccaaggagct gggactatag
37561  gtgtatatca ccaggtctgg ttaatttctg tatttttttgt agagacaggg ttttgccatg
37621  ctgcccaggc tggtcttaaa ctcctgggct caagcaatct acctacctca gcctcccaaa
37681  gtgctaagat tacagacagg tgtgagccac cacacccggc ctgattaaac aatttttaag
37741  aagcaaagga ataagaaaca caaaatggtt gatgttctaa ttcttgggct gggtaattac
37801  taggtttcat tatactatta agtgaagtaa aataaaaaag ggtcatgcat aaacaaatga
37861  tgacattgtt ttataaacct aaggattatg atgaaacact ctgtgcatac aagccctcaa
37921  caacaacaaa ggaataaaag gaaaagggag gaggagctag cagagaaaag agaaatgaca
37981  aataaagaat taccaagaat catttttttag gatttgatac agcaaggcta gtatttgcta
38041  atttaaacat catttagacc tttttggtat ggagaaattc cagtatctat cagaaaaata
38101  aacagactac aaaatgattt aaaagacaat agatttctta tacttactgc taaaagttta
38161  tctccaatct gaagtttgcc atccttatgt gctgcacctc cttcaattat tttggttaca
38221  tagatgctat tatccccagg aatatgctga tttccaacac ctccagcaat gctaaaccca
```

FIGURE 1-K

38281 agacctgttt ggaaaacagt tctagattct catccataaa tatgaactta agcccagaaa
38341 gaaatggttg aaaactacaa ggtaattatt ccctcacact aatatttttt aaaatagaaa
38401 agaatctatt tagttatcta ctagtatatg ttctgcaatc aagggacact tagacataca
38461 gctgctccta gagttacgaa tagattcctt ttatcccaca tgttccttct tagtatcttt
38521 tcgcaaatgc acttgttatt ccaaagtgac attgcgaaat tatgttttac ttttgtttcc
38581 atcataatta tgcttaccta cacttttaat attttcacaa gaacaaaatt attgagaaga
38641 aataacaaca ttgtctggtg atggtgggca gaaccaacag gctttaaaaa tgtgaatacc
38701 tttgttcttc ttcaatattc tatgaattaa caattttaag taaggaattt ctgtaattcc
38761 taaatcccat agcttcaata tcagataatt tctcagctct ccattccttt ttcatttgtt
38821 gacatcctcc ctcttccatg aagccacata ttgaaactac tatttccatt tatcctcaag
38881 ccatcaatgt tttaaaaatt cagcaatcac atcttaacag tcctttactg cagatgtaac
38941 tgctcataat acatattata aacacttctt cttttaactt agcctgtgta cctgcattta
39001 aagtattttg tagattatca caagttaata gcaatactaa acttcaaagt gttcaaggac
39061 acaaatattt cactctttta atgctagaag tcttcaatat aagaatactt aatacaaata
39121 gggacatatc cacttataaa acaaatagat attttgttcc ctacttttaa aaactcttaa
39181 aaacatgctg agcaaatatt gaatataaaa atcataaaaa acttataaat ttatatatat
39241 gctcactgca gtaaaatgta taaaagcata catcaataaa cgtaagaatt ttgggccagg
39301 tgcagtggct cacacttgta ataccagcac tctgcggggc tgaggtggga gaactgcttg
39361 agcccaggag tttgagacca gcctaggcaa tgcagggagg cctccgtctc tacaaaaaat
39421 aaaaaaaatc agcccagtgt ggtggtgtgt gcatgcagtc caagctactt gggagggtga
39481 ggtgggagga tcgcttgagc tcaggaggct gaggctgcag taagccatga ctgcaccact
39541 ggactccagc cagggtgaca gggcaaaacc ctgcctcaaa aaaacaaaca aaaaaacccc
39601 gcaaaaaaag gaacttttca aatgctgcaa tcttggtaga aatgtaaata ttctaaaatg
39661 ctacgtaaga caaaaatcag tagaaataac ggagaaatta aaatccactc aagttctgta
39721 gaatattact aatactgtac ttggaatgta tgtcacagat aaagttcata ggtatattta
39781 actcagagat ttcttaaaga tttatcttag tttgacttac cacatacctt taggaccttt
39841 aatgagcttt atttccatta ttttttctga cactggtttc cttcttttta catacaagcg
39901 tacaatagac cctgcttctt tcaacgcttc aactgctttg ctatgtgtta catcacgaac
39961 atctacttca tttactYgta atatacagtc attgaccctg agaaaacaat taaatattaa
40021 aaataaaaat taaatataat ttaaatataa attttccact ttaccaattt tttgttactt
40081 ctttttaag gtaaagagaa ttataaataa ttctggagta attccagaaa acataaatga
40141 agaaagtata tcaaaaacta atataaacaa atacaaacat ttcccaaggg ccagcaaaag
40201 gaacaaaaga aatagtgaaa ataatagatt atataaaaat gttaaataat aattacagct
40261 accatttgca tacttcctgc atgcttattc tgtgtaggca ttaagacaca ctttataaaa
40321 atagcaaaca tttatttagc actaaccaca tgccaggcac tttcttggta ttttaaccct
40381 catgacacct gtaagcttaa tatatatttt aatccctatt tcacagatgg agaaactgag
40441 gcacaaagaa tgtaaataac tttcctaagg ccacccagat aataagtgac agagctgtga
40501 ttcaaagata agaaaactga ggctcacatc acgagtttaa ggtcacagag atagtgtgaa
40561 aactgagata aagtaaaaat aattttctga gtgcctattc caacctatat attagacata
40621 aaacaattaa gttatttttt gaaatttata taaagattag gtcacttaat tcaaaggttt
40681 cgtaggtaaa gaacttataa tttgccatgg tgttatgttt tttaaaagtc tgatttgttc
40741 ctagtcttag ccttttccct agttctggtc tctttgattg catggtattc caccaggaaa
40801 aaaagatggt catctcaaat tgggtagaga gggtaaacaa aacataatta aaatattaaa
40861 actggtcctg acaagcttct atatctaaca gaatcaggaa gtaaatgtct acatttacat
40921 gaatggtcaa atgattaatt tttatatcta tttgatttca ttaatatacc acctgtcata
40981 cccagtgaga attataaagc ttagaaattc aaaaactgcg taaggcatca acgtccttaa
41041 aataagatgc tgtttttaac agactacata aactctttag tcctaaaact tcagggtatt
41101 tgaacaacaa acaaggaaat aggaaaaatg acataattta tgaccctcaa aaaatcaaag
41161 tttgcaaaat atatcttcca aaatggaaca ctaaatttaa acagagacaa atgttatttt
41221 catctagtca ataagaaaat atattattta aagtttgcat atagtctttg tggtgtggga
41281 ttcWacatgg atgtgtgatc ctctatcctc cgtatcttaa ataagtttat atacatgttt
41341 cccagccaga gaggaaatgg aaaatattct gtgagtgtac ttaaataagt ttatacacat
41401 gtttcccagM cagagaggaa atggaaaata ttctgtgagt gtatctgaaa acacaaaggg
41461 aaactcatga tttctaYaaa tatgctcaca aagatttaaa ccatcaacca aatYtgttat
41521 tcaaaatgtt ataatatcta agataaaact ggagtctaaa tgaaaacagg aacacaatgc
41581 ttagagcata attatctcat tcattgactt gaatgttgtt aaagcatcta ccctatgcta
41641 gttatcatga tcgtgtttgt tatttgtatt tatcccttat ccagtctgct tcaagaaaaa
41701 tacacaaagt ggattcagca tcacaataaa aagcgcaaaa cataccgcaa tcttccatct
41761 tgggcggctg ctccccctgt gataattttg gtaatgaaaa tacttgagtc atctccaatg
41821 tgtgggttgt ccgtacctcc tgcaatgctg aaaccaagcc ctgaatttcc ctgaggatag
41881 aagaaaaaaa attgagactg aatatgattt atctttatgg acaggttcct gtcattaaag
41941 gacgcagaaa ctgtactaac ataaggaaca caacaaataa aggtatatgg gtgagtaata
42001 taaagattct aaattaaatg gacttgttct ggaggaaagt cctttcagag gcattactgc
42061 tgaacaggtg aggttcaaga aaatacacca aggacttttg aagtatactc cacagttctt

FIGURE 1-L

```
42121  tctttaaaag tatgtgcaac ttgtcgttgg attagtcacc catctgcaga ggtgtcactg
42181  taagtctgtt tcttttagac tgcacctcag ccatgtttgg tagggcacat ctctcctaag
42241  agacaaatgc tgaagtgatc ttcccgctgt tatttctgag taggcaaaac ccatatactg
42301  ttttagcctt taaaaaaggg ggtggcgatg gtgcttttct tctgctaact tgacatcagc
42361  ttatttttaa atgaccaaat tgtggctgct aagtgatagc tgccatcagt ccctgacatg
42421  tttaaccaat atacccgatt agcaaaaggc attgtttcac actgactgaa agaccaacct
42481  gagaacttaa cattggtgaa ttttcttgac agcaaagtaa attttgccaa agcagctcta
42541  ctttttgctt gttttatgaa ctggagaaac aagatttctg aatatgttct tcacagaaag
42601  ggttcRactg ctacaaaaag cctttaaaaa taattgattt aaccaatggt attaaaaaat
42661  gtttatttct tgaagaatgt tctaactgaa tttcaggcaa ctcaagtaag caaatatatg
42721  tggacctatc atttaactgt tMgtattaaa cagtgtaaac acccactcta ccaccatttg
42781  tcttaaaatt ctattttaat aaaaaaatgt ttttcaatct aacttcatga attacctgaa
42841  tcttaagaac agtaccagtg gattatggga ggttaccaaa ttacttaggt actcaagttc
42901  ttcaaccagt gaaaaatata atgtatagtc ctaactaaac aaaaggaatt taaaaggccc
42961  tgggaaaaga taaaaaatca ataatataaa aagttaggag acatcacaaa ctactatatt
43021  aaagtatgta cagtacatta atactgtgct tacttgaata ttactccagt agccacccaa
43081  cttgagcaag tgatccgaat ttaagtagat ggcaataaag atatacatgt aagagactct
43141  ctcagattta tttttctcag attctattcc aatataaata aatctaatta tctaggattt
43201  cagtcttgag tatacacaca tttagttact taagaaatct gcttgtgtga aacactagat
43261  ttccataaaa tgctagccag ataaggtggc attactatta taacatacaa taaaaactga
43321  atgcagatat tctttttttct ttttttttaga cagagtctca ctctgtcacc caggctagag
43381  tgcgatctcg gctcactgca acctctgact cctgggttca agtgattctc ctgattcgag
43441  tagctgggat tacaggtctg cgttaccaca cctggctaat ttttgtattt ttagtagaga
43501  cggagttttg ccatgttggc caggctggtc ttgaacttct ggcctcaagt gattcacccg
43561  cctcggcctc ccaaagtgct gggattacag gcgtgagtgc agatattctt gagggaaaaa
43621  catatcttaa gaaaatttgc acatctgaat tgcaccttca agtcattttc tgattgggca
43681  agtataccat aattagtttt taatacagct ctgtctgtta gaagtatggt tattttgcta
43741  gaaaagaaaa ttgatataat gaccatgtta ctagtatact acttacttaa tacgaacaat
43801  acttcattat catatgcata taaacaatat aaacaaaatt aaaaagcaag aaaaatctga
43861  gaaactatca caatctagag gggtctaagg cggcaggata actaaacgta atgtagtttc
43921  ctggatgggc ctctgagaga gaataacata aaaagtaagg aaatagagac ttcagttaat
43981  aataatgtat gaatacaggt gtcttagttg aaataaatgt actctagtaa tagaagatgt
44041  taacaataga ggaggctgaa tccagagtaa atggaaactc ttgggactat ctttacaact
44101  tctttgtaaa tctaaaacta ttgtaaaata aaaagtttat taaaaaaata aggttactaa
44161  agaaaaggca aatgttttga aaatacctg gccaatttta tgaactggga acttagaaaa
44221  tatcatatag ttctgttttt gagagatctc ccttggattc tgtatgaaaa agcagtatat
44281  tacaaagttc tctagaacaa cagattaaga taatagtttt taccctttca agtgtgattt
44341  cttcatattc ataatctgca tctgtgccat taacctacag gggaaaagaa aagcagctca
44401  gaaacttcag agtgtaaaat ccaaaatggc aaggcaaaat ttttgtatga tttattaatg
44461  aaaacctgtt gaatgcgaca gtaatcaaac acaatagctt cattttaata attctgggga
44521  agattttcaa ggaatggaaa aatacaaatg tgattaatga gttgacaaca ttcgtggtat
44581  catgactttc aaatccagga taggtacatt cagcttttaa tagtttcaac ctccgatcct
44641  tctctgtctt gatccacaca tctaatcagt taccaactgt ttctgatttg aacaccttaa
44701  ttttttggaa aatcatcacc tcttcactct catttctaat acctttgttc aggccttgtt
44761  caacctcctc tggaatggaa tactgtcaat gcagtctgcc taatcgctta gattagtatt
44821  ccccactgcc caaatcactg ttttaaacag ctggatcttt tttgttttgt ttttttgaga
44881  tggagtctcg ctctgttgcc caggctggag tgcagtggtg gcgtgatctt ggttcactac
44941  aagctccgcc tcccgggttc acaccattct cctgcctcag cctcccgagt agctgggact
45001  acaggtgcct gcccaccacg cccggctaat tttttgtatt tttagtagag atggggtttc
45061  accgtgttag ccaggatggt ctcgatcttc tgacctcgtg atccacccgc ctcggcctcc
45121  caaagtgctg ggattacagg cgtgagccac cgtgcctggc aacagctgga tcttttttaaa
45181  gcccataaca tacttcctct aaaactttca caactctccc ttcctagaaa tatccaaact
45241  ctatgggttg gtaatcatgt cttcgaatga actgacccca atgtactact tcccttcatt
45301  cttcacctct tcataggcct tacataccaa ccctttctgt gaacgacttg ctctttccat
45361  tacatatcct gcaattttcc acttcgacat ctggaatttt gtgcctcctc ggtcacaaat
45421  ttcacttaca tgttttatgt caaaatcagt caccattcaa gatatcttgg tcacaattta
45481  tctccctcca atttatttcc tagtctctcc tgaaagaagt taactcctac agtaaaggcc
45541  actggtgaaa caagtccctg gaatcagatg aacatgggtt tgaatccttc ctttaaaatt
45601  tctcgtgata tgaggctggg caaaactcta tctaaatcaa atgtggactg tgataaattc
45661  tacttctgaa gatggctcac atcttccatt ccagacattc ttttgcaata taactaaatc
45721  accctaccat caaaaggagt ctctttcctc ttcaattaga tctaggcaag ccctgtgact
45781  agctttgact aaaataaatg tggcacaaat gatgctgtgt gatttctgag actaggtcat
45841  ctgatgcctg atagtattat ttgcctttgc ttggaatata ttttctttag aacccagcag
45901  ctataacatg agtaaaactc aaacagccta cagagagatc taaccaacag agaaccaaaa
```

FIGURE 1-M

```
45961  cccttggcca aaagccccag ctacactcct aaccagcagt cagcaccacc aactatcagc
46021  cacgtgaata aggctatttg gacctcccaa ccttgtcaag tgccctaacc cacaccatgt
46081  aaagcagagc catccaattc attatgagaa aaaatgaact gttgttcttt taagccccta
46141  agcttcaggg tagtttgtta tgcaatgata gataactgca agagaaataa caaaacctac
46201  ataagctact ttgctttctt aaccatacta taaaatgtag cccttaattc aggtctaata
46261  tagggaccag gagtttcttc cttgagaact tttgaagccc atctctacta ccctttggg
46321  ttcttaactc ttgtgattct cctagattca aaccctatta ctattttgtt ccctcttctt
46381  tctcatctat agacattttc ctcttagaga ttatatgtta taaaggcatt tatatagaga
46441  gagttgggat ttcactctgt tacccaggct ggaatgcagt ggcataatca tagctcgttg
46501  cagcctcaaa ctcctggact caagtgatcc tcctgcctta gcctcttgag tagctgagac
46561  tacaggcaca tgccaccatg cctggctgtt tacttttgtt tttcatagag atggggtcta
46621  tgttgaccag gagctcgctt gcttgctctc tctctctctc tcacacacac acacacacac
46681  acacacgtaa aatctcttta tctcagttta ggaggagtaa cttcaagagc tgtgttgatc
46741  atgtctgtgc atctagagta cgggaggctt acctacttga caggaagtcc tttgttacga
46801  gtaaaatttt tatttgcagc aatatcccct ttcttattct aaaggttatt tgcatcatct
46861  tgtttcttac tcaagacttg ctaaaagtct atcagtcttt atcgaataag cagttttgac
46921  attattgatg aatctattac atttcatggt attactttct gtttttaaca gttctttagt
46981  cctttgattt tctaaaaaag atttttgagt tgaatgctca atttatgctt actctattct
47041  atttttagaa acctcaaaat ttaccaattt tccttggcat atggattagg ccaaatccca
47101  caactgctag ctgcactgag ctccattttt gttcatttgt ttatttctaa ataatctaaa
47161  attttatttt cgaaaacatt tccaaatgag tgtgcttata tggccctttt ggcattaacg
47221  taaaattttg atactgtgac cttatgattt ctgccttaag gaatctgaga tttttcacct
47281  ttaagtttcc tcacaaggta ggtgctcagt aaatattcct ggtgcttaat ttaatgccaa
47341  ttcatgatgt aggtttctaa gcagcacata aaaggggctt tttaggtagc actgagtact
47401  ttactaaaaa tacaaaaatt agccaggtgt ggtggtgcac gtctttaatc ccagctactc
47461  agggcggggg ccaaggggtg gggtagggtg ggggctgaga caggagaagc acttgaaccc
47521  aggaggcgga ggttgcagtg agctgagatt gtgctactgt actccaacct gggcaacaaa
47581  cagagtgaga cactgtctca aataaataaa taaataaata aataaaataa aataaaataa
47641  aaagaactcg acccttttta caatagctaa aggaaaataa aatacttaag aatatactta
47701  accaaggagg tgaaagacct ctacaaagaa aactacaaaa cactgctgaa agaaatcaca
47761  gatgacacaa acaaaaacac atcccaagct catggacagg tagaatcaat actgtgaaaa
47821  tgactatact gccaaaagca atctacaaat tcaatgcaat tcccaccaaa ataccaccat
47881  cattcttcac agaactagag aaaacattcc taaaattcat atggaaccaa aaaagagcct
47941  ccacagacaa agccaaacta agcaaaagga acaaatctgg aggcatcaca ttacctgact
48001  ccaaactata ctatgaggcc atagtcatca aaacagcatg gtactggtat aaaaatggac
48061  acacagaatg aagcagaata gaaaacccag aaataaagcc aaatacttac agccacctga
48121  tctttgacaa agcaaacaaa aacataaagt ggggaaagga caccctattt gacaaatggt
48181  gctgggataa ttggctagcc acatgtagaa aatgaaactg gatcctcatc tctcacctta
48241  taaaaacatc aactcaaaat ggatcaaaaa cttaaatcta agacctgaaa ccattaaaat
48301  tttagaagat aacatcagaa aacccttcta aacattggct taggcaaaga cttcatgacc
48361  aagaacccaa aagcaaatcc aacaaaaaca aagatatgca gatgggattt aattaaacta
48421  agaagcttct gcacaacaaa ataatcagca gagttaacag gcaacctaca gagtgggaga
48481  aaatcttcac aatctatact tctgacaaag gactaatatc cagaatctac aaggaactca
48541  aacaaacaaa tcagcaagaa aaaaacaatc tcatccaaaa gtgggctaag gacatgaata
48601  gaaaattctc aaaggaagat acgcaaatgg ctgacaagca tatggaaaag tgctcaacat
48661  cactatcggg aaaatgaaaa tcaaaaccac aatgtgatat taccttaccc ctgcaagaat
48721  ggtcatcata aaaaaataaa aaaaaataga tgttggcgtg gatgcagtga aaaggaacac
48781  ttttacactg ttgatgggaa tgtaaactag tacaaccact atggaaaaca gtacagagag
48841  tccttaaaga actaaaagga gatctaccgt ttgatccagc aattctacta ctaggtatct
48901  acacagagga aaataagtca ttatatgaaa aagatacttg cacatgcatg tttatagcag
48961  cgcaattcac aattgcaaaa atatggaacc agcccaaata cccatttttc catgagtgga
49021  taaaaaatat atggtgtgca cacacacaca cacacacaca cacacacaca cacacaccat
49081  ggaatactac tcagccataa aaaggaatga aataatggca tttgcagcaa cctgaatgga
49141  actggagatt attattctaa gtgaagtaac tcaggaatgg aaaaccaaac acagtatgtt
49201  ctcactcata agtgggagct aagctatgag gatgcaaagg catgagaatg atacattggc
49261  atttggggac tcgggaaagc gtgggactgg ggtgagggat aaaagactac atactgggta
49321  cagtatacac tgctcagatg atgggtgcac caaaatctca gaaatcacca ctaacttatt
49381  catgtaacca aacaacaccc gctccctaaa aacttactga aattaaaaaa aaaatccaaa
49441  aaacaaaaaa taaaatcctg aagtttccca tcctttaaaa ttaaaaaaaa aaaagaaata
49501  aaaggcatcc agaggtaaac tatctctatt tgccagtgac atgatcttgt atatagaaaa
49561  ttcaaactac ataactacca gaggtaataa atgagttcag caaggttgta ggataaaaga
49621  taaacaccta aaagtcaatt atattcccat atgctagcaa taaaaactcc aaaaataaac
49681  aaaacaattc catttataac agcatcgtaa gaataaaata ctggctgggc atgatggctc
49741  acccctgaaa tcccagcact ttgggaggct gaggcgggct ggatgacttg agctcaggag
```

FIGURE 1-N

```
49801  ttcaagatca acctaggcaa catggtaaaa ccccatctct accaaaaata caaaaaaaaa
49861  aaaaaaatag ccagtcatgg tggcacgtgt ccgtgttccc aactacatgg gaggctgagg
49921  tgggaggatc atcttgggtc tgggaggtgg aggttgcagt gagctgagat tgtgccaatg
49981  cactccagcc tgggtgacag agtgaaactg tctcaagaaa aaaaaaaaaa aaaaaaaaga
50041  gagagaatac atagaagtga atttaacaaa agaactccaa gacttgtaca cagaaaacta
50101  aagaacatta ttgaaaggaa ttaaagaaat atacccatat atttacagtt aactgatttt
50161  ctatatggtt gccaagacca ttcaatgcgg aaagaatagt cttttcaaca aacagtgatg
50221  ctatccacat acaaaagaat gaagttggac ctctacttca cattatacaa aaatcaactc
50281  aaaatgaatc aaagaccaaa atctaaaagc aaaaccataa aactcttaga agaaaactta
50341  actctaagtc tttgtgatgt tggattatgc aatggcttct tacatacaat acaaaaagtt
50401  acaaaagaaa aaatataaaa ataatggttt ttagactcaa tcattatcac gaaattgcgg
50461  cattttctaa agttttatcc acttaaactt caacaaaaat gtagcagtat atttttaccc
50521  catactgtca ccagcagaat gtactgacaa actttttgct cttttgataa cataatgttt
50581  aagatgtatt aattgtttta atttgcattt atgttattaa ctcacctttt tctacttttt
50641  taccttaaca aattatgttg agatattcac atatcataaa atctaccttt ttaaagtatg
50701  caattcagta gtttctagtt tattcacaga gttgtgccat catcattact ttctaattcc
50761  agaacatcct tattacccca taaacaaacc ctgtacctat tagcagtcac tccccattct
50821  gccctacacc caccacctgg caactgctaa tctactttca gtaactatgg atttatctat
50881  tctggaactt tcatataata aaattatgca acaagtggac ttttgtccct gatcaattta
50941  aaattacaaa gttataacta attattttcg tatatatgaa tgcacaaaaa aaacctggtt
51001  actaaaaatc tagtcattca attagacttc taaaatgatt taaatatatg gcactgtgta
51061  agcaaccaat ggtcagagac aacctttttgt gaaagttagc catcataatt tctgaattaa
51121  attacgaatt tctatttcaa aatttctgct ttacaaaagg aaaaaaaaat gaccaaagga
51181  ggaccatact acttaattta ctatagcaat tacagctata gtatatagtg atagaacatc
51241  ttaagagaaa tgtattagag aaaacaatgc catggtcttc gcatttgtat caaaaaacgg
51301  aacaggaaag gcagaattac ttcaatgtgg ggaggaaatg gaacttacgt aagttggtgt
51361  ttccaagcta tctgtgttga ccagtactgg gggaggattt gcctttaaga agaaattgta
51421  aatgtgttaa caagaaaata aaacattaca tgttcatgtt cacaatgtta gaggcatagg
51481  aaagacttgc tatttcttaa gttataatct tatatgagct ttttatatgt taaataatat
51541  atcctttctc cagcaccaaa cagaaaaaat tatgtataag cagttaaaac agatttgctt
51601  ttaagaaata aattgctatt tctgctaaag cagccatgat tttcttaagt tttacacaga
51661  tcaccatagc cttctcataa aacccatgaa tattctttag tgtaaccaat aactcttcct
51721  aggttgctta gaaatcattg ttttctcttc ccaaagatac agtaaagatt tcaatcattt
51781  catcacccag gagaaaattc ttccccacag tgaactgttt ttaaacaagg aaaaaaaaaa
51841  gttcatcaca gaaagaaatc aattgtaaac ttttagaaac taaaaacaca ctcttaatgt
51901  aaaaacccga ttattttcag ctgcaatgac tgctttccag ttattctaga ccactcacag
51961  aattaacagt gatttattac agcaagaaaa ataagcagaa taacacacac ctagttaagt
52021  ttggattgaa aaaaaatctt ggttttctgg ttaggttaga aattattatg tatatgaaac
52081  taaagaataa aatgcaataa caaatgaaaa atataataaa aatctgatgt aattcccctc
52141  aaaattcttt cttctgcata tttaaagtta tgaaagtacc aagtcacata acactttcag
52201  agaataaaaa agatgactat cctttaaaca aactttagaa gatatgctta cacacattaa
52261  tggttctact tcaattacct gaggaattat acataaaagc tgtataaagt aacactgtaa
52321  cacaaaataa ccattttctt ccaaatgtac aggcttttaa taaaggggaa tattcaaact
52381  tcaaataaca cattattttc agcacagtac cacagctaaa tgatatgaca aaatgcttaa
52441  agacaaccta gtttcatggt taaaaaatta aacagtatgc atttgggcta catcaagcag
52501  taagttaaaa atgagtagct aatacttcat tatatataaa aataaaatta aatataatac
52561  tttatgttta gtcccaccat taaataatag tctacctttg aataattcgg tgaaagtcca
52621  ctcaacaaat aaaaattgta cttataaaac tgctatttaa gcaaagctac ccaacttaca
52681  gacatttgag attttatgaa tatttcaaat tacttatttc ccaatggtat attgagatta
52741  ctcaaactgg tattttacgt tgctaaaaaa attaataata caacttcact ttattctgct
52801  taagcaatta cagaagaaaa gggcaagaaa gaataggata tttaaagtta caaaagaaat
52861  tacattttaa aaacactatg gcaagaaagc cttttggtcc ctctgccgat tttttaccaa
52921  acatgcagtg gaaaaagttg gtttttatca aaatctaatt tataccataa ttattatgtt
52981  agctgtttat cagtaccaac tcatcatcat tttatagaag ttacctttg attctattca
53041  tttagaaagt ctcgctaata ttgaaatata tataaacgaa atgatataag gtgttatatt
53101  tacttcaaaa taaaacaatg gaaggttggg agtaggaggg gttatgaatc aaatagaagt
53161  acacgagttg attgctgcag ctggtgatgg gtatacatag agttcattgt actattcttt
53221  ctacatttga atatttctga aattttgcat aataaaaagt aaaccaaaaa atcactttga
53281  gccaagtact cagttttaaa gaggttaata gctaaagtag ttgtgaaata ctctgggtca
53341  gggttgctat taagatactg gcattcagcc aagtgcagtg gctcatgcct gtaatcctag
53401  cactgtggga ggccaaggct ggtggatcac ttgaaaccag gagttcaaaa ccagtctggc
53461  caacatgggt gagcaaaatt atgtttctta agccttatta ccgatgcttg ggtcaaatta
53521  tttaggggct ttttagttcc tccttctgga atatccttgc ccttttactg cctctacccc
53581  caatctctaa gtgaattata gagcttcact gttgtccaag ttggaactat acacaatggt
```

FIGURE 1-O

```
53641  atataaattt tagtaacaga ataagcattt ctatagcttc ctatgatttg caacatattt
53701  tcttttatat tacttctcta gagcctcaca cagtcccccc gggtaaagac aggtattgct
53761  atcttttcac agatggagaa aatgaatttt aataataact gtgatttacc tatactcaaa
53821  tgtcactgag tggtagaggg aggtatagag agacctagaa ccaggtctga gattacaggg
53881  tcacttgaac caaggtgcaa tgagggcttt ttcttcatct gaaataattt cacgtagact
53941  ctttccaagt cttttagttg tagttgtttg attttccatg ttacctgttc tactactcgc
54001  tgtcttggtt aaggatccca aagtcttttt ttttttttt ttttttgaga tggagtcttg
54061  ctctgtcacc cagggtagga ctttttaaaa ctcaagttta ggccgggtgc ggtggctcac
54121  gcctgtaatc ccagcactct gggaggccaa ggcgggtgga tcacgaggtc aggagatcga
54181  gaccatcctg gccccgtctc tactaaaaat acaaaaaaat tagctgggct tcgtggcagg
54241  cgcctgtagt cccagctact tgggaggctg aggcaggaga atggtgtgaa cccaggaggt
54301  ggagcttgca gtgagccgag acagcaccac tgcactccag cctgggcaac agagcgagac
54361  tctgtctcaa aaaaaaaaaa aaaaaaaaaa aagtcaagtt tattggctgg gcacatggct
54421  cacccaaatt ttaccattct atttcatcaa caaaggatct tgtagtgatt gctgctgtaa
54481  cagactgaag cagccttaca aagtttcaga gggcaatatt acatttttaa aaatcgtgtt
54541  aaaatacatt ctaagatcca attattttaa aataggtcaa aatagtttag aaaatccact
54601  tttacaatct gcattgtatt tctccttctc agcattgatc aaagtttcat tatgttattt
54661  gtctagtgtt tctttctcct ctactcaatt atgaacaagt gctttacaag tgcttttcca
54721  atgcaaaact cacaaactcc agagaattct tcttagatcc taatttgaga cacattgcat
54781  tcaaagttaa atgctatgac tgaggtgatc actgcaggtt aaaatgattg taaaaacaaa
54841  tggggtggat tctggaaaga aggtagagta ggaagcacca caaatctgtt tctccacctt
54901  gaacacaatt ttagtggcag catgtgtctg atgtagatgt tttggaactt ggagtctact
54961  gaaggctggc aacttccagg tgaactgtag ttaatttcag tgacctgcag gcttagcaca
55021  gcagcagctc cccacactct atccctcagt caactgtgta ccgcctgcac aagccacggt
55081  gggcagagaa aattacattc tccaaatact ggagatctgt gctctgaatg ctgcttctca
55141  tcacaaaggt gcaaacaggc cctggatact gctgttccac ctcctcctca ttgttggaaa
55201  ccccacccca acctcccagg ctgaactgac ttccaggaga ttgaaagagc tggaggcctc
55261  cccttcctac acaccaattc aattttcttt ttcccctttt gggagccaga cattgaagac
55321  taggacatta aaaaccaacc acatgtatgg aggaaattag aaagtcactg tgcatgccca
55381  ggggaaggca taggctcaga aaatatctaa gaagacccta cgtttacacc tgaggctgat
55441  ctttgatagc ctacaacaat cagaaaaaca ataacaacaa aaaaagaaaa ccctggagaa
55501  gaaggagaat ctgattttca gagttaccaa attattaaat tcaagtgtgc tgtttaaata
55561  aacaagaata agtggcttat taaaaaaaat ataaatcaat ataaactgtt ccttaaaaaa
55621  aaataaggct gggtgtggtg gcttacgcct gtaatcctag cactttggga ggctgaggcg
55681  ggcagatcac aaggtcagga gtttgagacc agcctggcca acatactgaa accctgtctc
55741  tactaaaaat acaaaaaatt agctgggtgt ggtggcaggt gcctataatc ccagctactt
55801  ggaaggctga ggcagaagaa tcacttgaac ctaggaggtg gaggttgcaa tgagccgaga
55861  tgacaccact gtactctagc ctgggcaaca gagcaagact ctgtcaaaaa aaacaaaaca
55921  aaaccaaaaa accaaaaaaa cccacaaaaa cctgatagca aatctactag acaaagactt
55981  tattaaagaa ttaaaagatg gccgggcaca gtggctcaca cctgtaatcc cagcactttg
56041  ggagactgag gtgggtggat cactaggtca ggagatcgag accatcctgg ctaacacagt
56101  gaaaccccgt ctctactaaa actacaaaaa attagccggg cgtggcggca ggtgcctgta
56161  gtcccagcta ctcgggaggc tgaggcagga gaatggcgtg aacccgggag gcggagcttg
56221  cagtcagccg agatggcacc actgcactcc agcctgggtg acagagccag actctgtctc
56281  aagaaaaaaa aaagaattaa aagatgtgaa caaaagcaag aaagtgctgt atgaacgaaa
56341  cggaaatatc aatgaagaga aataaaaatt ataaaattca ggaaatgaga agtacaataa
56401  cagaaaattc actggagaga ttcaaaagca tatctgagca ggtaaaaaaa gtagtgaaca
56461  tgagatagga caagggaaag tactgagtct gaagaacaga aataaaagag attcaagaaa
56521  agtgaacaga acctaaggga cccgtgagac atcatcaagc agaccaacta atgcattgtg
56581  ggagttgcat gagaaatgac agataaaaga gtagaaaaaa tatttgaata atagccaaaa
56641  cttctcagat ttcatgaaac acatgaatat aaacatccaa gaagctcaat aaacaataat
56701  gaaatccaac agactcacac tgagacacat tatactagaa ctgtcaaagg ccaaaaacaa
56761  agggagattt ttgaaagcaa caagagaagt gacttgtcac atacaacaaa tcctcaatga
56821  gattatcagc agacttcaca tcagacactc tggaggtcat atggcagtgg gtggaaatac
56881  tcactgctaa aataagacga aaaaaaccca aacctgtcaa ctaagaatcc tatatccagc
56941  aagacagtcc ctcaaaatta agggggaaat taagatgttc tctgatgaac aaaagctgag
57001  ggagtttgtt atcactagaa ctgccctgaa agatgtgcta aaggtagcag ttcaggttga
57061  aatgaaagaa aactagacag caactcaaag tcatatgaag aaataaagat ctcagtaaag
57121  gtaaatacat aggtaataat aaacactagt tagtaatatt gtaacaatgg ttatgtaaat
57181  ctgcttttgg ttttccacat gattgaagag accattacat tttcaaattt aaaaaaaaaa
57241  aaaacttagc ctagccaggc atggtggctc acacctgtaa tcccagcact tcaggtggcc
57301  gaggcaggca gagatggctt gagcccagga gttcaagacc agcctgtgga acatggtgaa
57361  accccatctc tacaaacaaa acaaaacagc aacaacaaca aaaattaacc aggtgtggta
57421  ggcacatgcc tgtagtcatg gctattcagg agactgaggc gggagaatcg attgggccca
```

FIGURE 1-P

```
57481  agaggttgag gctgcagtga gctgtgatgg caccactgca ctccagcctg ggctacagaa
57541  tgagaccctg cctataaata aataaataaa taagcctaat attaataaac aaagtcaata
57601  ttggtaaata acgaaaatca attattagtt taaaagctaa cattataact ttgatttgta
57661  acttcatatt ttgtctccta cataatttaa gaaacgaacg cattaaaaat tactagtttc
57721  tgtttttggg catacaatat atgaagatgt aattctgtga catcaacaac tgaaaggggt
57781  tgggacagag cagttaaagg ggcagaggtt ttgtatatta ttgcagttaa gcttgtacaa
57841  attgagatta gaagtgtcta ggatgttaaa tgtaatcccc atggtaacca cacaaaatat
57901  aactaaagaa tagacacaaa ggaaacaaga aagttaaatg tttcactaca aaaaattaat
57961  caaagaccaa agaagacagt aatgcaggaa atgaggaaca aaaaagctac aaggcatata
58021  tataaagaaa acaaatagca aaatgacaaa agtaagtctt tccttaccaa taattacttt
58081  aaatgtaaat aaactcttca atcaaaagac agaaattggc agaataaaaa ttttaaaatg
58141  ttccaaccac aagctgtaca caagagactc actgtagatc cagagacaca aatatgctga
58201  aactgaagga cagaaagggg tatttcatgc aacagtaacc aaaggagagc aggagtggct
58261  gtactcataa cagacaaaat agactttaaa taaaaaaaag gttatgagac aacaaaggta
58321  ttatacatta ataaaaggtt caatatagga atgtaacaat tacaaaaatt aacgcaccta
58381  atagcagacc atcaaaatgt taagtagcaa aaatgagaca gaattgaaga aagaaatggt
58441  tctacaataa tagctggaga cttcaatacc acattctcca taatgggcag aacaaccaga
58501  catatgataa gtaaggaaat agaggagatg aacaaacaca atataccaaa gagacacaga
58561  actctaacaa taacagaaca cacattcttc tcaagtgcac atgggaatag aaggaaacta
58621  tgtcaaccta agaaaaacca tatacaaaac acacacagtg aacatcatac tcagtggtga
58681  aagactgaaa gcttttcctc taagataagg aagaaggcaa gtatgtctgc tttcaccact
58741  tgtactcaac atgaccacta gctgaatagt tgaagttgtt gtcaaagcaa ttaggcaaga
58801  aaaagaaata aaagacatcc aaattagaaa ggaagaagca aaattacttg ttcacaaatg
58861  atatgatctt atatgtaaaa caccctaaag attctacaca aaaactgtta gaattattaa
58921  accaattaag caaagtagca ggatacaaag tcaatacaca aaaatcagtt gtatttcttc
58981  taacactgaa caatctaaaa tggaaattaa gaaaacaatt ctgtttatta tagcatcgaa
59041  aagaacaaat tttcagaaca ctgagcctcc taaatgaaga attaacttca tcaagaaagt
59101  aaaaaacttg ggcaatgaaa actataaaac atgtatgaaa gaaattaaga agacataaat
59161  aaatgggaag ggatctgtgg tcatagattg gaagacttac tattgcaaaa atgtcaatat
59221  tacccaaagc aatctataga cttaatgcaa ttcctatcaa aatcccagta gggttttcaa
59281  agaaatagaa taacccatcc taaaagtcac atagaatttc acggtaccct gaaagccaaa
59341  atggtaatga aaaagaaaaa caaaggtggc gggctaacac ttcctgattc caaaacttac
59401  tacaaagtta cagtaacaga aacagtctgg tactggcatg cagacagaca tacagaaggg
59461  aataaaacag aatccagaaa taaatgccat atacaattat caacctacaa tggaYcatga
59521  tctaaatgta aaacctaaaa cttaaaactg ttagaagaaa acacaggcta aaagcgagac
59581  actggaattg tcaatgattt cttggatatg acacaaaggt acagacatgt cttgtctgta
59641  atctctgaca agacatgaga cccagaatac acagaggaac tcctaaaact cgacgataaa
59701  accaaacacc ctaattaaaa aatggtcaag gaactcatac agacattttc ccaaagaaga
59761  cacacacatg gacaataagc acatgaacag atgtgtcaca aatgcaaRtc aaaactacaa
59821  tgagatgtca cctcacaccc actagcctgg ctactatgaa gaaaacagaa aataaaaagt
59881  gttggtgagg atgtggagaa attggaatcc ttgtgcactt tggtggaaat ataaaattct
59941  acaactggct ggatgcagtg gctcatgcct gtaatcccag cactttggga ggccgaggca
60001  ggcggactac ctgaagtcgg gagtttgagg ccagtctgac caacatggag aaacctcgtc
60061  tctacttaaa aaaaaaaaaa aaaaaaaaaa ttagccgggc gtggtggcac atgcctgtaa
60121  tcccagctac tcgggaggct gaggcaggag aatcacctga acccgggagg tggaggctgt
60181  ggtgagccga gattgtgcca ttgcactcca gcctgggtaa caagagcaaa actccatttc
60241  aaaaaaaaaa aaaaaaaaaa aaagSccaga ggcagtatga tgctctgctc tggacatacc
60301  tctaaggtaa ttactctcac attatgcggg agctatttgt taaacttgtc tcttgcacaa
60361  acacctagta ggccccctaga ggacagagac gatgttctat tcatcttcaa agcacatatt
60421  caaatatcaa aaaagaccat gcacaaaaaa ttagctctta aagcattttc aacaatactt
60481  taattacatg atagcttttc agaactgata gaaataaagg tttaaaacat ctagttttaa
60541  agcagagtat ttactctagg gtgcaaataa gcctctggat ttaataggct agtatcacag
60601  agattatgtg tttacactcc cagtaagaag aactagtaac tgtcacctac tctgtactca
60661  gtttctatgt ggagaaactg aggctctcag aagttgagta atttccacac catcacacgt
60721  agaaacaggt gaagctagga agtggtggag tcgggtagga ctataaactc cacgttcttt
60781  ctgcaatatt aagcagccat taaatattac ctttatctgt gccactctgt ataataagca
60841  taattctgat ttgtagaaga ctttcataaa gtacaaacaa tatgatcaat gtgaaagtac
60901  tctgaaaagt ataaaagtgt tctacaaatc atgaaagact atatactttt taaaaagttt
60961  tcatctatgt atctttctaa tttgcctgac tctcaaactc attttaaagg agtcaggttg
61021  gcgttacccc ccattttaat agatgaaggg ggtataaaac tcagagaggg taactagctt
61081  gagtggcaaa gacagactag attctaaatc ttttcttctg tttttatttc taatacatcc
61141  taacgcatct aaatgtaaag tagtggatct tttaagaata catattcact taatatgttg
61201  aaattgggtt atatgttagt atgtatttta aattttactt ggggacggat attttagtcc
61261  attattttaa ttttataatg tacacattgt acttcactaa ttaggaacac acttattctg
```

FIGURE 1-Q

```
61321  gaaaatgagg tgcactcatt ggcttctcac ataacacaac aaaaatggta aaactatctt
61381  tcatgaactt ccgtagtgta tttaaaacct aaagtgaagc tatgcagaat aaataggtct
61441  tttttgtgtg tatcctcaat aaatatcctc atttaaaaaa acaaaacaaa acttgttcct
61501  tagtttgctg tctcagagaa atatgagata acactaaaaa tatgggcggc taagattatg
61561  tgccagtacc tgacgtgcat taagcatcct ctccatcact gtaaagcgcc tcaagttagt
61621  actcctgtct atcttcagtg ttacagatga gtaaatggaa gtgcgctgtg gttaagtgaa
61681  gtgtgcagtt acacaggtta agtggtagga ctggaattcc aaccccaggt cattctgact
61741  ttaaaacgtg tatttttata agactatgca agatctcaac aattttcaaa tgcagtggga
61801  tgctaagtta gggttgcttg gagtaaataa atgagagaat tctgaatagg aaaagacagg
61861  ctcccaaatg aatataaaga gactggtaca aattctcaaa ggattccaga aaactatttt
61921  aacaccaaaa aattgactga agtgagaaga gtaaaataaa ctttcaattc tattaaaaaa
61981  aaaaaaaact gtactaataa actctatata tttggaaaga tgttacatta ttatgatgcc
62041  ataataaccc ttaaactgtt tgctcagaat tatattttat ttataaattc aatttttttc
62101  ctgcaatata gagagaatat ccatttggat tcatctattc attgttggtt ttatgactta
62161  atttttaatt attttcataa tcaaaaatta tatagtagca gtattattat aatgattaca
62221  attcagtctt attacagtaa ggctgaagtc attgtaacac tgtaatttcc cctagaatcc
62281  ttggttgtat ggatgttccc actgacttca ctttttagga gaggggaagg gcactgagat
62341  gtgaataatc acagctgata tatcattttt tccccaatgg agcttatttt gaaactgttt
62401  acacagcaaa catatgttta cccagatttt aggctacaga tcaataatgg caaaaaaaaa
62461  tctatgggtt catttattta cactttaaat gagatcgtga tctttcacac caaagatctt
62521  tattattatc tttaatacca ctaccattat tttgactttg gataactgtt ggaagggcag
62581  ttaatgtctc aagccacccc ttggaagtag taaagctttg aaaaactaaa aatgattaga
62641  ctctctcaca tttggctatt atttttaaac tgttacacat aatttttaa aagtaaaatt
62701  ttacaacaga tttacagaaa agttgtgaag atagcacaga gcattcccat atactcccag
62761  atatgttggc tattattttt acctaataaa tgtgaagtgt gttatctaat caaataacat
62821  aagaatgaat atttccaaag aagataaata tataacagga aacaaagtgt taaataaaat
62881  gatacattgg tacaagatca ggagtacaat gatttttact aactttagaa gaataaagtg
62941  ttctaaagta aaaatgttta aaaactgttt taagcagatt tctaagtaca aaggcatgaa
63001  aacaaatcta aggttcaata ttatgaagtt cattgtgtcc catttcacag acaacctaat
63061  aaaatggcaa atcctgactc cacttatatt aaatcccaaa tgtttaagtc cctaaaacta
63121  ctaactgaag cccaaagtaa attaccaaat tacaaaaaac agctcaaaca gttttaaaaa
63181  agaacacaaa tgaaacatta aaagcaaatg aaatttttat aaagaaaaaa gaaactaata
63241  aaaacctgtg gtatggtggg taggatgaca gtattctcag cagggactgg caggacaggg
63301  atcacaggga cagtgggagg agagggaaga acagcttctg ttggctaaaa atcatcaatg
63361  acaaaaaatc aggaaaaaca ccaaaataaa agggaaacac ataacaaaaa gctgtgccat
63421  aaagataaat tatagttttc agaaataact atgcaacagg aaaaacatta atacattaca
63481  ctaatataaa tgagaaaata ttctctagaa attgttttag aaagttatac ccccataata
63541  attttaatga ttctgtaata aaatgaatgc ctactcttta gatgcttaaa tcataataat
63601  aaaagattat catttacaat attccttcct tctctgggag ataagacact tgcttcatag
63661  acagactgca aaatattttt tctagcactc agaatacatc tcactacttt cctctatgca
63721  ttcagaaaac taaaagtaat tttattagcc caatcaggca gacaaagaga gtttacaaat
63781  ccgtgatact aacttctgtt atcaagaaca gtaataactt gcctcaacat gcctgtaatt
63841  aattcaacta acattgattg tctattacat gccagacata tctggcaaaa atgaaaaaac
63901  aacgtagaga gaacctggta taaattaaac taaaaagttc actgttttat ctggcttgtt
63961  atttttaaaa ggatgaaact ggaacacagg aaagttgttt agtaataaga cccatttgct
64021  atataaataa aatatcttat atatgtaaga aaaacactaa aatcaagcag atgaggacag
64081  cctgccttaa caggccttaa aacaggaaga acaaattgcc agactaaaaa aagggcttat
64141  cttcattcat aaatgcaaaa taaaatggca atgatctttt ctgcctacca aattagtcaa
64201  attaagattt attaacaatg tatggagtct cttcaacaat ggtgctgaga caactggata
64261  tccacatgca aaagatgaag gtctatagac ccccatctca cattatatat aaaaattaac
64321  tcaaaattga ttaacaacct aaatatgaga tttgaaaaca taaaactctt acaagagaac
64381  atagggttaa tctttgtgac cgtggatttg gtaatggaac cttagaaatc ataaatgatg
64441  aaagaaaaaa atcaattaga ctctaacaga attaaaacct tttgtgcacc aaagggcatt
64501  accaagcaag tgaaaagaca gcctacacaa tgggagaagg tatttgcaaa tcatacacct
64561  gataagggtt taatatccag aacatttaaa gactcttaca acgcaacaac acaaagagaa
64621  acaacccaat taatgaatgt gtgaagagct tgaataattt ctgcaaagaa ggtatacaag
64681  tggccaatta gcacacgaaa agatgctcga catcattagt ccttagggta atacaaataa
64741  aagctataat gagagattac ttcaccacta caagggaagt gtctaattaa aacaaaacaa
64801  aaaacaaagt aacaagtggt ggcaaggatg tggagaaact ggaactctgg tacaatgctg
64861  gtgggaaatg taaaatggta cagcttctga ggaaacttttt agtggtttct taaaaaccaa
64921  ccatagaatc agcatctgat ctagcaattg ggtaggtaaa tatgcactgg gtaggcatat
64981  tcccaaaagt gagagcaagg acttggacac ttgtatgcca atgttcaatg cagcatcaca
65041  cacaacagtc aaaaggcgga aagaaaccac gtgtctatca ggagatgaac ggatacacaa
65101  aacgtgataa tatacacaca atgggtatga ttttttttttt ttttttgaga tggagtctcg
```

FIGURE 1-R

65161 ctctgttgcc caggctggag tgcagtggtg caatctcagc tcactgtaac gtccgcctcc
65221 cgggttcaag caattctctg cctcagcctc ccaagtagct gggatgacag gcacctgcca
65281 ccatgctcag ctatttttt ttttttttt ttttttgtat ttttagtaga gacagggttt
65341 taccatcttg gccaggctgg tcttgaactc ctgaccttgt gatccaccca ccttggcctt
65401 ccaaagtgtt gagattacag gtgtgagcca ctgcacccag ccctaggaat aaaattctta
65461 atcatgctag aagatggatg agcctttaaa acattaagtg aaattagcca gacacaaaag
65521 gataaatatt gcatgattcc acttaaaaga ctagtaagtt atacatattt taccaataaa
65581 aaatattcca aaaaagcttt taaaaatgca tgaaaactgg tcctctcaca ctgctgttgg
65641 atatacaaat tcacacaaac tttagggaaa gtaatttggc agtaaatatc tagagcttta
65701 aaaatgtcta aacttggcca ggcacagtgg ttcacacctg taatctcagc actttgggag
65761 gccgatgcgg gtggatcgcg aggtcaagag ttcaagacca gcctggccaa catggtgaaa
65821 ctccatctct actaaaaaaa aaaaatacaa aaattagccg ggtgtggtgg cccatgcgtg
65881 taatcccagc tacttgggag gctgagacag gagaattgct agagcctgga aggtggaggt
65941 tgcggtgagc ctagattgcg ccattgcact ccagcctggg caacagagtg agactctgtc
66001 tccccctcct caaaaaaaag tctaaactct ttgacttagt aattctagaa atctacccta
66061 aggaaataat tttaaaagcc tatgttttaa ggtactccca taaggtgttc taaggtttta
66121 agatagtctt ttgaacactg cttctatttg gtaaatcgtg gaatatttga aatacccaaa
66181 ggtgtcagaa ttgttaagta aaaaatgaaa actctatgag attgaggctg ggcatggtgg
66241 ctcatgcctg taatcctagc actttcggag gacgcggcag gcagatcacc tgagtttagg
66301 agttcgagac cagcctggtc aacatggcaa aaccccgtct ctactaaaaa aaaaaaaaaa
66361 agaaaaaaaa tacaaagatt agctaggtgc ggtggtgcat gcctgtagtc ccagctactc
66421 gggagggtga gtcaggagaa ctgcttgaac ccagcaggtg gaagttgcag tgagcagtga
66481 gatcacgcca ctgtactcca gcctgggcaa cagagcaaga ctatctcaaa aaaagaaaga
66541 aagaaagaaa agaaaattat atgagattat catacagaaa tttaaaatta tcagtacgaa
66601 cagtttataa attatatggg acattgctta tgacaaaaca ttaagtaaaa aagggaagat
66661 gaaacacata aaaaatggaa agaaatgtct gaaaaattaa tctctggggt gggatcatga
66721 gtgcctttct ctatcgctgc ttctttattc tgaattatac ttttccagtt gttgacaagg
66781 aatacacagt ggcgttatta atataaatgc aaggggaagg atgggagatc cattgtgaat
66841 attaacagaa aaacaattat tatgatcttt aagtgtcatt ctctagaaaa caaactgaaa
66901 atgtatttca cactctggac tcctgttacc tggatttatt atttcctgta gaaagagatt
66961 tttaaagcat tggattttgt gttttctctt tcttttcaat gttcccagtc tgggatgatg
67021 aaaagttctg aagattgaga gtgatgatac ttgcacaaca ctgtgaatgt acttaatgcc
67081 actgaattgt acattaacaa atggctaaaa tggtaaactt gatgtaaact ttaacacaat
67141 aaaaaagtat aYtacgaata aagcctccgt gggtttagct cgggactcca atcaccattt
67201 atatattgtt cctggagaaa aaaaattctg aattccacat tatgagctca aaattaataa
67261 caatcgtaac tgatactcac tgcaatattt acatggcctt ctaagaatgt ttgaaaaccc
67321 ataggttctg acttctgatt tatgaatgca taagaaaagc gtggtggttt tgtaagaact
67381 gtgcttggat caagtggata tacacatgga aaaatcaaac aaatactgat tcctaccctt
67441 aaactataca taagattaat ttcagctggt ttgctgacat aattgagaga agtaaacaat
67501 ataaggcttc tagaggatga tatggtagag tatcatcaca accttggtgc agatgacctc
67561 ctgaatagga aagtgcatgt tgctgaagct cacgcagggt gaggaaagca tctgtaggca
67621 gggagtggaa gggaggcaat agcctggcaa agggtgctgg agcctcagca tggtgacgag
67681 ggtgtctaca tggacaaagc cgtggcggca gcctgtgtgt ggttggaacc caagtgagac
67741 agatagcatc tatgaggcag tagctgcagt ctggcataag gttttgtgct gtcagtggac
67801 taaatgggat cttctggagc ctgagcagcg tgagaagggc gtccaactag ggcagcaatg
67861 atgacaggag attgcttata tacagggcat tctggtccaa taaataaata catgaaggat
67921 actgacagca ggtttccttg cgtcagaaaa ggcagttata aatatgaaaa gatagaaaat
67981 tagcatgaga cctatggtgt tacactaaaa ttggaggtat tggtataaac agttttcaat
68041 atatatagat aaatacaaaa ataaatatag acataaatgt gtgtgtatat atgtgcatag
68101 gtctgctaac tatgaggacc tggatgcagc taacagctca acagcacgaa tacacctggc
68161 atccaaatct tgatttctaa atactcattc tgcaatttaa aacaacacaa catggccggg
68221 cacggaggct catgcctgta atcccaacac tttgggaggc ggaggtgggc ggatcatctg
68281 aggtcaggag tttgagacca gcctggccaa catgatgaaa ttctgtctct aYtaaaaata
68341 caaaaaatta gctggacatg gtggcgggca cctgtaatct cagctacttg ggaggctgag
68401 ccaagagaat cgctttgaac ccaggaggca gaggttgcag tgagccaaga tcacgccatt
68461 ccactccggc ctgggcaata agagcgaaac tccgtctcgg ggggaaaaaa accacaacac
68521 aaaaccaaac aaaaccaaac catgggtctt tgaagaaatg gctgactcca gtgctagagc
68581 tcagcaagca caagatgtgt gtagacatct tgctgtacca gaagataagg gaagtactca
68641 aagaggatga caaaggacac agaagttggc ttgaagggct gccactggcc agactgggaa
68701 caatctgtaa ctcaatgata acaatggcat aacatccact gaataaaata ggaatctggg
68761 agcccatact gatataaata aatacatcaa taaataaatg aagataagag cattttctta
68821 caacaaaatg ctaactaaaa atgtaatgat ggaattataa aatcatcatt tggcagccat
68881 catagtaata gagtaattca ttcaagaaat attaatgaat attaaaactt cctgttgcca
68941 agtgaaacgt ccattccagt ggatgaatgg gatatactca tacaaaagtg tctcctttca

FIGURE 1-S

```
69001   aagcattaat gacagtggag aaatctgcag acaatatctg aatcaagtga tcaaagttaa
69061   caccaccagt aacgggaaaa atcaaaatca tatatcatgt gataaaatgt aatgagagta
69121   caatacttcc aggatatttc taccaaaggt atataacttg aatctaatca ttaaacatca
69181   gacaaaccca aattgagtga tacattacaa agtaactagc tgcagtcatc aaaagtgtca
69241   aagtcatgaa agtctaacgt tatttttaat tcaatttat tctagtttaa aaacaacttt
69301   actgaggtag aagtgatata aaaaactgca catatttaag gtatagaatt tgatgagttt
69361   gcacataggc atacatctat gaagccactg ccataatcaa agtaacaaac gtatctatca
69421   cctctaaaag cttccttgtg cctctgttgt tttttgtttt ggtttggttc ttttaaggac
69481   actaaagttt atttttttaag tgtttgaaaa tgtcacaaat gtaatatgtt ttcttattaa
69541   cagtctccag ccctaacaat aggataatag ctgattaata atctactttg atagccaaaa
69601   taagacaaaa gtacatgctt atgtcaaatt tcaaagcaaa gcacaaaaat tttctccact
69661   gggcaggtgc tctccaaaac cctttatgat aaaccatgaa tcctttaatg ttacatttta
69721   atcccacttg taaccctctg tctctgagcc tcttctagaa tcaccttctt tagacgagga
69781   tgtgaatacc atttaaaaaa attcccagtt tacagcatct taacccatcc ccctctctac
69841   agggcactga ccaagttcct tgaagaacat aggcccctgc ctcttcctta aggccaacac
69901   cgctgagaaa gtcaacgacg catcaacaca cacaattccc taatcacaac tgataaccaa
69961   ttctggtaac aaacaattac agaacaacaa agtaattcat ttagttcact tgaaacaaaa
70021   atgcagctct aaacaaataa taatcaaatg aaggctgact gatgaggaac agcaaagata
70081   tcagatctga tcacaggatg ctttgggcaa aggttttaga gcagggtaaa actccgggga
70141   gctgtataag attctaaatt agaaaatgtc aaacttaaga ttctaacata aataatgaag
70201   cagcaatata cagtaaccat tatattttgg ttattacaga attcagtatg gcattttaaa
70261   cttgaaaaat cttatataac tatttcatta gacagtgact gttatgccat catccttgtt
70321   ttgctccatt tgtatcactc aataagatga gatctaaagt ttttctgttg ttgttttaaa
70381   aaaaatcttg accatgtgac tagcaacggt ctcagttgtt tctatccaca gcaacatcat
70441   ggaatcatta gcctctgagg ctgccttcag agacttgttt ccatggcaac agaggctatg
70501   atattaacac cgtcacaagg aaattgtaaa tttctacaaa attggatata ttgtacatat
70561   tttaaagtac cttttaaaaa ccagtatcaa atggatggct atctgaacaa atgtttaaag
70621   tgatttgttt ttagagatta attttctcta tagctttata ttttaaatct aatttaagat
70681   cttaaaatta taaaagcagc acttaaaaca gattttcata aagtggtctg aacactcact
70741   acatacagcc atttgcatct gttctgcaca tccacattta tgaccatgta ccgccctccc
70801   tcactttgat tcattgaaaa ataaatgtat attgaatgag cctactactt ttattaaagc
70861   cttttctgct taaagcactc aagtggtttt cattatgtgc atatgaaccc tgattaatac
70921   aggaagagac agaagggaca attaggagtt tactgtaata atccagctga aaggttattg
70981   gtggcttaaa ctgaatggtg gtaatggaca tatgagcagt caaatcaaga cacattttga
71041   atgtagagtt ggtatatctt ggtgactgat gagaagtctg gagtgaggaa ggcctattaa
71101   cactcacaat agccactaaa aatgatcYtt tatacaagag ttcgactcag gtataaaaga
71161   gttgcaaggc tatctaaata tatttaaaaa cattttctca ctaatttatt aaaatgactt
71221   tcatagcata agatgcaact ccaaagacga taaatgaaca tactatctga tagctaataa
71281   tgcttatatt ttacattaac acgttatatg aaggggtaa actgacctaa gaatatacac
71341   aaagattcac tactgtatct ttttaattca acttttattt taagttcagg ggtacatgtg
71401   caagcttgtt acacaggtaa acttgtgtca ctgggatttg ttgtacagat tatttcatca
71461   cccacatatt aagaatagta cccatttagt tatttttccg gatcttctcc ctcctcgtac
71521   tctccacccc tcacaccggc cccagtgtgt tgttccgctc tatgcgtcca cctgttctca
71581   tcacttagct cccacttata agcgagaaca tgcggtattt gatttttttgt tcctgcatta
71641   gtttgctaac gatgacgccc tccagctcca cccatgtttc tgcaaaggac aagatctcat
71701   tcatttttat tgttgcatag tattccgtgg tgtgtatgtg ccacattttc tttatccagt
71761   ctaccattga tgggcattta ggttgactgc atgtctttgc tactgtgaat agtgctgcaa
71821   tgaacctaag cgttcatgtc tttataacag aacaatttat attcctttga gtatacaccc
71881   agtaatagga ttgctgggtt gaattgtatt tctgttttta ggtctctgag gaaccaccac
71941   attgtcttcc ataatggttg aactaattta cactctcatc aacagtgtat aagtgttcct
72001   ttttctccac aaacttgcca gcatctcttg ttttgatttt ttttaaagta gccattctga
72061   atggtgtgag atggtatctc gtggctttga tttgcatttc tctaatgatc agtgacgcta
72121   agctcttttt catgactgtt ggccactttt gaacagtgtc tgtttacgtc ctttactctc
72181   ttttaatagg actgtttttc tcttgtaaat ttcttgaagt tccttataga cgccggatat
72241   tagacctttt tcagatgcac agtttgcaaa aattttcttc cattctgtag gttatctgtt
72301   cactctgttg acagcttcct ttgctgtgca gaagctcttt aattagatcc cacttgtcaa
72361   tgtttgcttt tgttgcaatt gcatggtgtc ttcatgatga acccttttgcc tgttcctata
72421   tccagaatgg tactgcctag gttgtcttct taggttgtct tatagctttg ggtttcacat
72481   ttaagtcttt aatctatctt gagttaattt ttgtatgcac acttccatat ttttagagaa
72541   aatgagatat atgaattagc aaatcttttg gtaccaatat cttagatggc aaggaaaagt
72601   atcaactgtc gtgtcactgt gggttagtaa ggatcttatt ttggttttgt atggcataat
72661   taatattcct aatctaagtt ctttggttct ttgtttaaag atgtattatg ctgtaatgtg
72721   gcatgtttct ctgattaaat gtcaaagaag ctatttaaaa gtaagagaga aactaaaaaa
72781   aactgcaagc cttttgaatg tttatagaag gataaaatat agcacagtag atttataaaa
```

FIGURE 1-T

72841 actctgctta ggaatcactg tactcaacgg gggcccagta ccttcagaaa aagctgattt
72901 ttatggttta aatctgtttt tttattcaaa tgtcctttgg tcctgtcact gcaaacaggt
72961 atctttcaca taagccggta tagtaaaaaa ccaacaatat gttatttgaa ccagcttcat
73021 atgacttcca tgtcataact atctcacaag caaacaatta taaaactaaa cacattttta
73081 aacaaatcac taagggtatt aaaatcatta aatttaaaac aaaacaaggt aaataggaaa
73141 acagatctag caattaattc cattacagaa acatgagtac aagtgtgtta gttcacacag
73201 acagatcatc tagttcatga atctatctct acctgtgaga ctcacaatat tcagttaagg
73261 tttctcatcc ctccattatg aaaagacagt attctatcag cattcagatg acaatcttgt
73321 cctagtaaag gatgagccaa acacgaaaca tttcaaagta aacagacact aaaacttcaa
73381 tatatggtta cataattgac aaatggggat aaaattacca aatgcaataa tgtgcttttt
73441 tacctgattt ttatcagaat gtggctagag aaataagcaa aaccacagac ttcatgacta
73501 ataacctaaa atgggccctt aatgcttccc agtaatcaga tttgccttgt attattcacc
73561 ctctacaagt ctgtaaggat tcaggtccct gttatcactc tctcaagttg ccttcttcta
73621 tttcccccac acaaatacac acattttcta ttgtttcttt cctttatttt tcctccagaa
73681 taccatatac actcatacgt cacatatcaa cattttggtt aatgatggtc cacatgtata
73741 aaattataat ggagctgaaa aatttctact gcctagtgat gtcatagcca ttgtaatgct
73801 gcagtataat gtattacatt tgtggtgacg ctggcgtaaa caaatatacc tcactgccag
73861 ttgtacagaa gtataatata tataattatg tacggtatat aatacttcag aatgataatt
73921 aacaactatt ttactggttt gtatatttat tatactatac tttttatctt ctgtataatt
73981 atatacaaaa ataaatatac accgtattcc cattaactga catgtgactg cattttggat
74041 aattatattt attatctgtc ttcttctact tagagcagat ttctgttctg ttcactccta
74101 tatgcccagt gcctagaaaa gtgcacacag gaagtattta atatatattt gttgaatgaa
74161 taaaaccatg cgactgtcaa aagaaatcaa cttaatcagc actctaaaat gacttgggca
74221 atagattggg taggtacagt cttccttccc ctacccactt tcatgcaata tacatagtag
74281 acattgctaa aaggttattt ctgtgtttct gggcagccac caccaatcag aaggtgttca
74341 agatgtttta cagtttgata accttagaac agaagtttca aaatgtaaaa aattgccagt
74401 ttcatgttgt tctctatcta gcaataatga attgcctaat aatgtcccat ttgaaaataa
74461 acatgtcttg gtcaagcata tttgcttaaa atattttcat cgcttttcta ttcaaactgc
74521 attaggaagt acaacttctt actgagggct catatttaaa atccatggtg cccttcctct
74581 gaggttccac taatttcagc ttgggtcctt tgttccccta caggtagcta cttctgtaag
74641 ttagtctcga ttttaccttg gcattctcat tttgcctttc cagttctcca atactcatct
74701 aaccaatccc atgtagtgta tttttattgt tagttgctgt ggtttctgtt tccctggctg
74761 caccctatta aagttaccaa ccctctagct aaaacatggt tgcttctccg tatctctaga
74821 ttcctcattt aaaggtcaga atattttgtt gaaatgaaaa tgatggtggt ggcggtggta
74881 gtggtggata aaagggtagt tgataaagat gatgatgacc tcactgatgt aattaatgta
74941 ttataataag tgctttgcac atattgtctc atttaattct cattaaattt tgaaaaagat
75001 aacactttca tccccatttt ctacagatga ggagacagac tccgcatgga ggtaatttgc
75061 ccaaggtggc acaggcgata tggcagaata taaatttaaa tccaggtcat catcctctaa
75121 agatggtaca cttacccagt actctttact tatctcttta aaatagttca aaaagagatt
75181 ttgtctttat tgttaccact acattttact gcaaagacaa tatatattac ctgtaaaaag
75241 aaatacatac tgcatcaatc cttagatcag gggttagcaa acatttttgg aaatgacaaa
75301 aaaagaagcg ttatggacca tgcagtaatt gttgcaacta tccaacactg ctattgtagt
75361 tcaaaagtag tcatacacaa tacacaaata ggcatggctg tgttccaata aaactttatt
75421 tacaagccag gcatggtggc tcatgcctgt aataatccca acactttggg aggccgaggc
75481 aggcagatca catgagggca gtagttcaag accagcctgg ccaacatggc gaaacccagt
75541 cagtcaggta tggtggtgca catctgtaat cccagctact taggaggctg gggcataaga
75601 atcacttgaa cccgggaggt ggaaattgca gtgggctgag atcgcgccac tgcactccag
75661 cctgggcgac acagtgagac tctgtctcaa aaacaaaaca aaacatttat ttacaaaaac
75721 aggtggcagg ctacatttgg cctacagatt gtagcttgct gactcctgct ttagctgctt
75781 ttaaaaaagg agaaagaaaa ggctaaataa gtcaatgaag aaaatagaaa atggctttaa
75841 gacacctaag ttaactttta ctaactataa atggtttaat tctgttgcaa ttcttagtct
75901 acaataaaaa taagccaatt ctctctcccc aaaaaccatg gaagcgacaa tttaaaaaac
75961 aaggaaacaa gatatacata tggaaatttc aagtttcacc tacagttctt taaaatatct
76021 ttaattaaaa agtctatata ataaagctgt gctcttacaa ctgaaaaaat tttgcctcta
76081 tactgaatat agccttccag tttggaaaaa aagtgattaa aatgatatga tttttattct
76141 cttttctttg ttctcccttc aaaatcctgc tatggtactt agtttggcta aagcaattaa
76201 ttcttaatat tgatatgcat cttgtttgtt ttctttctat cttacaaaat actttccaac
76261 tgatttcttt ccaaatttat aacccttga tacatcagat catcctacca agattagtca
76321 gatttatata cagataaatg agggtgggat ctgttttaat ttaaatgcca gtttattgct
76381 cttgaacgtg gctgttaaaa aaaaattaac aaataattaa acaccaRgtt ataatcccca
76441 attcttttc tcagaacaac aaaacaacac aaattgtaaa gcaatcattt ggttccactc
76501 tgtcacagct gtaacttctt cacatttcaa agctgctcca gtcaagaagt cagagttttg
76561 atttcacaga tactgaaaat actgttcaga aatgctacca

FIGURE 2-A

>7:10710001-10808300

```
1       cagattcaag ttgccctgaa tatactccaa caagcagtag ttacaagtgg attttaaaaa
61      aattattatt atacttaaag ttctagggta catgtacaca acgtgcaggt ttgttacata
121     ggtatacatg tgccatgttg gtttgctgca cccatcaact cgtcatttac attaggtatt
181     tctcccaaca ctatcccttc Yccagcaccc caccccccta caggcctcag tgtgagatgt
241     tccccatcct gtgtccatgt gttctccttg ttcaactccc acttatgact gagaacaggt
301     ggtgtttggt tttctgtcct tgtgatagtt tgctgagaat gatggttcca acttcatcca
361     tgtccctgca aaggacatga actcatcctt ttttatgact gcatagtatt ccacggtgta
421     tatgtgccac attttcttta tccagtctat tattgatgga catttgggtt ggctccaagt
481     ctttgctatt gtgaatagtg ctgcaataaa cacacatgtg catgtgtctt tatagaatga
541     tttataatcc tttgggtata tacccagtaa tgggattgct gtgtcaaatg gtatttctag
601     ttctagatcc ttgaagaatc gccacactgt cttccacaat ggttgaacta gtttacagtc
661     tcaccaacag tataaaagcg ttcctatttc tccacatcct ctccagcatc tgttgtttcc
721     tgacttttt aatgattgcc attgtaactg gcatgaggtt gtatctcctt gtggttttga
781     ttttcatttc tctgatgacc agtgatgatg agcatttttt catgtgtctg ttggctgcat
841     aaatgtcttc ttttgagaag tgtctgttca tatcctttgc ccacttttg atgggttttt
901     tttttcttgt aaatttaagt tctttgtaga ttctggatat ttgcccattg tcagatggaa
961     agattgaaaa attttctccc attttgtaga ttgcctgttc actctgatga tagtttcttt
1021    ttctgtgcag aagttcttta gtttaattat atcccatttg tcaattttgg cttttgttgc
1081    cattgctttt ggtgttttag tcatgaagtc tttgcccatg cctgggtcct gaatggtatt
1141    gcccatgttt tcttctaagg tttttatggt tttaggtgtt acatttaagt ctttaatcca
1201    tcttgagtta atttttgtgt aagtgtaagg acagggtcca gtatcagctt tctacatatg
1261    gctagccagt tttcccagca acatttatta aatagggaat ccttttccca ttgcttgttt
1321    ttgtcaggtt tgtcaaagat cagatggttg tagatgtatg gtgttatttc tgaggcttct
1381    gttctgttcc atttgtctat ttatctgttt tggtaccagt accaagctgt tttggttact
1441    gtaggcttgt ggtatagtct gaagtcaggt atgcctccat gtttgttctt tttgcttagg
1501    attgtcttgg ctatgcgggc tgtttttggt gccatatgaa atttaaagta gctttttcca
1561    attctgtgaa aaaaatcagt ggtagcttta tggggatagc attgaatcta taaattactt
1621    tgggcagtat ggccattttc acaatattga ttcttcctat ccatgagcat ggaatgttct
1681    tccatttgtt tgtgtcctct tttattttgt tgagcagtgg tttgtagttc tccatgaaga
1741    ggtccttcac atcccttgta agttgcattc ctaggtattt tattctcttt gtagtaattg
1801    tgaacgggag ttcacttgtg atttggatct ctgtttttct gttcttggtg tataggaatg
1861    cttgtcattt ttgcaccttg attttgtatc ctgagacttt gctgaagttg cttgtcagca
1921    aacttcagcc caaaaggaga ttttgggctg agatgatggg gttttctaaa tatacaatca
1981    tgtcatctgc aaacaggggc aatttgactt cctcttttcc taattgaata ccttttattg
2041    ctttctcttg cctgattgcc ctggccagaa ctttcaatac tatgttgaat aggagaggtg
2101    agagagggca tccttgtgtt gtgcagcttt tcaaagggaa tgcttccagt ttttgcctat
2161    tcagtatgat attgggtgtg agtttgtcat aaatggcttt tattattttg agatatgttc
2221    catcaatacc tagtttactg agaattttta gcatgaaagg ctgttgaatt ttgtcagcct
2281    tttctgcatc tattgagata atcatgtggt ttttgccatt ggttctgttt atgtgatgga
2341    ttatgtttat tgatttgtgt atgttgaacc agccttgcat cccagggatg aagccaactt
2401    gatcatggtg gataagcttt tcaatgtgct gctggattca gtttgccagt atattattga
2461    ggatttttgc atcaatgttc attggggata ttggcctgca attcttttg ttgtatctct
2521    gccaggcttt ggtatcagga tgatgctggc ctcataaaat gagttaggga ggattccctc
2581    tttttctatt gattggaata gtttcagaag gaatggtacc agctcctctt tgtacctctg
2641    gtagaatttg gctgtgaatc tgtctggtcc tggacttttt tttggttggt agcttattaa
2701    ttattgtgtc aatttcagaa cctgttattg gtctattcag agattcaact tcttcctggt
2761    ttagtcttgg gaggggtatg tttccaggaa tgtatccatt tcttctagat tttctagttt
2821    atttgcacag aggtgtttat agtattctct gatggtagtt tatatttctg tgggatcagc
2881    ggtgatatcc cctttataat tttttattgc atctatttga ttcttccctc ttttcttctt
2941    tattaatctt gctagcagtc tatctatttt gttgatcttt taaaaaaaca gctcctggat
3001    tcattgattt ttttcttgaa gagtttttg tgtctctatc ttcttcagtt ctgctctgat
3061    cttagttatt tcttgtcttc tgctagcttt tgagtttgtt tgctcttgct tctctagttc
3121    ttttaattgt gatgttaggg tgtcgatatt agatcttttc tgctttctct tgtgggcatt
3181    tagggttaca aatttccctc tacacactgc tttaaatgtg tcccagagat tctggtacat
3241    tgtgtctttg ttcttattgg tttcaaagaa catctttatt tctgccttca ttttgttatt
3301    tacccagtag tcattcaggg gcaagttgtt cagtctccat gtagttgttt tgagtgagtt
3361    tcttaatcct gagttctaat ttgattgcag tgtggtctga gagacagttt gttgtgattt
3421    ctattctttt acatttgctg aggagtgttt tacttgcaat tatgtggtca atttagaata
3481    agtgcaatgt ggtgctgaga agaatgtata ttctgttgat ttggggtggg gagttctgta
3541    gatgtgtttt aggtttgctt ggtgcagagc tgagtccaag tcctggttat ccttgttaat
3601    tttctgtctc cttgatctgt ttaatattga cagtgcggtg ttaaagtctt gcattattat
3661    tgtgtaggac tctaagtctc tttgtaggtc tctaaggact tgctttatga atctgggtgc
```

FIGURE 2-B

```
3721  tcctgtattg tgtgcatata tatttaggac agttagctct tcttgttgaa ttgatccctt
3781  taccgttatg taatggcctt ctttgtctct tttgatcttt gttggtttaa agtctgtctt
3841  atgagagaca aggattgcaa cccctgcttt ttttttctct tccatttgct tggtagatct
3901  tcctccatcc ctttatttta agcctatgtg tgtctttgca aatgagatgg gtctcctgaa
3961  tacagcacac tgataggtct tgactcttta tccaattttc cagtctgtgt tttttaattg
4021  gggcatttag cccatttaca tgttggctaa tattgttatg tgtgaatgtg atcctgtcat
4081  tatgatgcta gctggttatt tcacctgtta gttgatgcag tttcttcata ctgtcgatgg
4141  tgtttaccat ttggcatgtt tttgcagtgg cttgtatcag ttattccttt ccatgtttag
4201  tgcttccttc aggagctctg gtaatgcagg cctggtggtg acaaaatctc tcagcatttg
4261  cttctctgta aataatttta tttctccttc acttatgaag cttagtttgg ctggatatga
4321  aattctgggt tgaaaattct tttctttaag aatgttgact attggccccc actatcttct
4381  ggcttgtagt gtttctgttg agagatccac tcttggtctg atgggcttcc ctttgtgggt
4441  aaccagacct ttctctctgg ctgcccttaa catttttttc ctttatttca accttggtga
4501  atctgacaat tatgtgtctt ggggttgctc ttctcaagga gtatctttgt ggtgttctct
4561  gtatttcctg aatttgaatg ttggcctgtc ttgccaggtt ggggaagatc tcctgaataa
4621  aattctgaag agtgtttttct aacttggttc cattctcccc atcactctca aatacaccaa
4681  tcaaacatag atttggtctt ttcacatagt tcttggaggc tttgttcatt tcttttcact
4741  ctttttttctc taatcttgtc ttctctattt taaccatttg atctacaatc gctgatatcc
4801  ttccttctgc ttgatcgaat tggttattga agcttgtgta tgcttcacgc agttcttgtg
4861  ctgtggtttt cagctccatt aggtcattta agctcttctc tacactggtt attctagtta
4921  gccattcatc taacctttttt tctaggtttc taccttcttt gcgatgggtt agaacatgct
4981  cctttagctc ggagaagttt gttattaccg accttctgaa gccttttttct gtcaactctc
5041  aaactcattc tccatccagt tttataccct tgctggagag gagctgtgtt cctttggagg
5101  agaagaggtg ttctggcttt tggaattttc agccttttttg ctgtggtttc tcctcatctt
5161  agtggtttta tcttcctttg tctttgatgt tggtgaccta cggatgtggt tttggtgtgg
5221  atgtcctttt tgttgatgtt gatgctattc ctttctgttc ttattttacc ttctaacaga
5281  caggcccctc agctgcaggt ctgctggagt ttgctggagg tcctctccag accttgtttg
5341  cctgggtatc accagcggag gctgcagaac agcaaatatt gctgactgat ctttcctcta
5401  gaagctttgt cccagagggg cacccacctg tatgaggtgt ctatcggccc ctactgggag
5461  atgtctccca gtcaggctac tgggggggtca gggacccact tgaggaggcc gtctgtctgt
5521  taatggatct ctaacaccat gctgggggaa ccactgctct cttcagagct gtcaggcagg
5581  atgtttaagt ctgtagaagt tgtctgctgc tgccttttgt tcagatatgc cctgccctaa
5641  gaggtggaat ctagagaggt agtaggcatt gttgagctgt ggtgggctct gcctagttcg
5701  agcctccctg cctctttgtt tacactgtga gcatagaact gcctactcaa gtctcagcaa
5761  tggtggactc ccttcccccc accatgctcc agcgtcccag gttgatctca gactgctgca
5821  ctagcaacga gcaaggctct gtgtgtgttg gccttgccga gccaggcacg ggagggaatc
5881  ccctggtctg cctgctgtga agactgtggg aaaagtgcag tatttgggca ggagtgtaca
5941  gttcctccag gtacagtcat tcacagcttc ccttgtttag aaaagggaaa tcctctgacc
6001  ccttgagctt tctgggtgag gtgacacccc accctgcttt ggcttgtcct ccatgggctg
6061  cacccactgt ccaaccagtt ccagtgggat gaaccaggta cctcagttgg aaatgcagaa
6121  atcacccatc ttctgcatcg atcttgcagg gagctgtaga ccggagctgt ttctattggg
6181  ccatcttgga agtgcccccc aaggagcttt attgagagac agaacagctc tcagtgaaga
6241  ggagacccaa aatacctagc tcccatcctc aggcaggtag tcctgaaatg tgtctgagtc
6301  tggctgagtc tgaggttttt atgggctcag aatgtagaaa gtgcatgctg attggtctat
6361  gggtgggccc agaaaaagca ccatctgatt ggccRaaagg catcaaagaa cttctcactc
6421  ctggtcatgg actctaccca gaacaggtag cctggcccac aggcttcagg ccatccctgg
6481  cttgaaggtt ggtcctcact ggggacctac cctttcccgc gtaggaacct gtctgcctcc
6541  caccaccatc aaaaccttga aagcccatct aattggtgga atctaaatca tatctataat
6601  cctagtttta agtgagtctg agaaacgtag tttttattct aacatgatct gtaatacaga
6661  ggaagaggtg agaatgaaag tcacctgcta gtggacaata tctagcccaa ttttatgctt
6721  tgtctactca gcaaccagat acacccttct atccgtaaac ctccccacaa tcacaatagc
6781  aataaaacca ctaccaccaa acataaggta actcttccta atatagatga aaacactgtc
6841  tccttcaggg tcagaatttg tccttgcccc ttcgaacata ctgggatctg attccattta
6901  tagggatgca gtgagagaga tgggcatgag gagaatcagc atccctttgc aaagtaatga
6961  actcagttga gtccattcat tacagtgtct tcaacaaggc aggtagcttc accaagaaag
7021  aagggtctaa ttctactaag aaaatgatgc tctgagattt tgaggcttgg ttttcaattc
7081  tcaaggtacc attctatccc aataaatatc ctaaattctt acaaagttgc ctaaaacctg
7141  caacttgtgt attcagattt atcaattgtt aatgatatca gtcctatgat gacatgggat
7201  gaaattaatt tagagtagta gttttcattg gagggctgag tgatggtggt gccactgttg
7261  ttggtagtga tttggctccc cagtggatat ttggcaatgt atgtccattt tgagttgtca
7321  caatttgggt tagggagagc tgttagtggc atctagtgga tagagtccag aggtactgag
7381  atatcctaca atgcatagga cagccttcct cagcctttg ccacagggcc aaagaattat
7441  ctggcccatc aataaggaca aagataagaa atcacagaag ctctctgtgt gttggaacaa
7501  gttttgagtc aagcagttaa cgtcagctgg atgtagtcac tcacacctgt acccccagca
```

FIGURE 2-C

```
7561   ctttgagagg ctgaggcagg aggatcagtt gagctcagga gttcaagccc agccttggca
7621   acgtagcgag gtcccatgtc acatttagaa atttttttaa aaagttaaaa ccctgggctt
7681   gtcatcattg ttctctaatt tttcaaaggc tgtcagaact gccacctcag cccaccccag
7741   tccatttatt gcttatttgc attttagtgt tctatcataa caaacaattg ctaacagatc
7801   ttgtgttcaa caggcacctt atttcagttg atgacatatg ataatttaag taaatgtatt
7861   accactgcat gtcttggact accaatatct tattttccac tggtaatcaa gatatcattg
7921   cctttaaatc caaactgggt caaataataa atatcaaatt cccagtgttg aagttttatt
7981   tctttatttt ttgaacatct ttatatacca actattgtat cagacaaaat tagtatagca
8041   aatagaaaac actacatata tttcatgtag aaaaatattt tggtatgtat aaaatcattt
8101   aaaagactga aggagtgaaa gtcagagatt caaggaaatt ttcagtttca agatcatata
8161   actgtagcta tgatctggag gtcagaaaac tagtgcagct gttgctgcta accctctcca
8221   caactgctcc tccatcttgc ccttatacct ttgaattaag ctagtgaata tggattctat
8281   tcttatagat gattgagctt cttttctgct gttacagcta cacgaacatg gcctctgcct
8341   cccttacatt ttctacatat caagatgccc atgggatacc taatacatct gagttaagta
8401   attgtttaaa gtttttaaaa atttggtttc tccaatgttg ttagttttat ttaaatattt
8461   aatgattctt ggttgtctat atatatttgt gttcatagta tataatacgt agtgttgatg
8521   gccagaattg atgtactttg tccaaataag aagccctRtt cacctggcag agaataaatg
8581   caagttttca ttataggtat actggttctg atggttgtgt gtttgtttgt agggttgagg
8641   aagttggtgg tatagggagg ttacatgtga gtaggctctt ctttatgatt tatgagtagt
8701   ttctatgctg gatcttccta tttcagctat taagacaact gaaagtcagt gaggttacta
8761   tcagtctcct tcctataaac ctgttttcca ttctggtaat ctctgaagtt agacttcccc
8821   cttagtagag ctgtatttga agttttattc aggagagaaa tgctattttc agctcctctg
8881   acaacagaat tactatgtgg ctcttttcct tctaatgaat tatccagtct attgtccaag
8941   tctcagccct tctcacaacc ctcattgtct cttacctaag cgttatgcag agaatgtatt
9001   ggaaaaactc actttagtga ttcttctctt tgatgattca tcacctggat gaatttatta
9061   gctccttagt ctgttggatt ttgttggcaa tgtggaccta cttccagaac tgtcaaaatt
9121   cctccatagc cctcaaaagt ctgctggtga cattactctc aggtttatag caatgctatg
9181   attttatgga tttatttcta tttgtatatt tcttctgtga tttctgtgta ttttggggag
9241   ggtagtgagg taattcacaa gttcaattag tttaattagt tgtgaaccag cactattatc
9301   cttagattta cagcaaacta attttacttt cttggtttct taaaaaatta atcaggatct
9361   gcagtttctc tctcaacaaa ataagacaat tcatttttt cttcatttat cttattttac
9421   tttccctaRg cacttcatct ctttaccccc tacatcttgg atttggttac cgtaatttag
9481   aattatgacc cctctagttt gcatatcctt gcttaataaa actgtattaa atcccacaat
9541   ctcagtgttg acaattattc acaattaatt ctaacttatt ttctttttct cttacatcat
9601   tttttcattg catgtcctct aatttcttgg atttaaaaga ctttttttctt agagtaattt
9661   tggaaatagg gtacactgtt atctgcttct caaaacttta agtaaccaga caggttatta
9721   atattatttc cagattctgg gtgatgaatc agaggctcaa tgttgccttt ttttagggta
9781   acacaaattc agtggccaag cctagaccMa aattcggatc tccagactct taagtttgct
9841   gtcttttgta tacatatttg tttctttttc ttacagtagg tgtccacaac aagctattta
9901   ttttactatc ttttgttgga gtagtttgaa agaccatttc tgaatataga actctgcttt
9961   tcataattat tactgatgct tgtgcagtca tatttagttt gcttaattag acagaattgt
10021  ggacattaag aatattattt ggcaaaatcc tacattgagg cagtttgtgc tYgctgtttt
10081  ttggaaatga taattattat ataattcttc ttctttgtgt agtcatgcat aaatatcttg
10141  aatttgaata cctggcaact agaataaaag agaaacatga tggattacat aacataaata
10201  ttctaccaag aagatggcag agcattagga gactcaaaat tttctgagtt aaggctgggt
10261  gcagtggctc acacttgtaa tcccagtact ttgagaggct gaggtgggcg atcagttgag
10321  gtcaggagtt gaaggccagc ctggccaaca tggtgaaacc ctgtctctac taaaattata
10381  aaacttagcc agggatggta gcaggcatct gtaatcccag ttaccaggga ggctgaggca
10441  ggagaatcac ttgaacctgg gaggcggagg ttgcagtgag ctgagattgc tgccactaca
10501  cactccagcc Rgggcaacaa agcaatactc catcccccta cagctccccc aaaaaattct
10561  gagttaagaa tgccaagagt aatttatcac atacgcctta aagatgttca gcagaatatt
10621  attggagatg ctattatagc cagtttacaa ccaataaaga caaattcttt atatgacatc
10681  aaaatgtaat ttccttttata aatatgggtg caaaaatcaa caaaatacta gcaaactgaa
10741  tctagcatct ataaaaagat ttatacaaca tgactaagta gaatttatac tagaagtgca
10801  aggttgatta aacaaataaa aatcaagcaa tatagtaata tactatattg aaaaaataaa
10861  gtacaaaaat cacatgatca tctcaataga cagagaaaga gtatttaata aaattcgaag
10921  ctctttcatg ataaaaacag ttggcaaact aggaacagga aagttcctcc acctgaaaaa
10981  gggcatcttt gaaaaactat agataacata agactaagtg gaagactgaa agttttcttc
11041  ctaaagtcag aaacaagaca aagaagtctt tccctcacca catttattca acaatgttct
11101  gcaggttcta ggcagagcaa ttaggcaaga aaaagcaata acaagcattc acaaaaggga
11161  gggaaaagtt caagcccctt catttaatat catttcctgg aagttgcacg tactgtttct
11221  gctgacatgc aattggccag aacttaactg cataactcca agcaacatca agaaagttgg
11281  tatattatgt ctttattctt tgaaaactta tgcccagtta aaatttggag cttccatttt
11341  tgaagaaaaa ggagataatg ggtattggga aacaaccagc agttctttcc atgtttattg
```

FIGURE 2-D

```
11401  aatgcaggtc aaaaaaagga attaggtaat ttttttcttt tctttttacc ctacttttct
11461  tcttcctttt tccctccctc tcttccttcc ctcctttctt ccttctttcc ttctcttact
11521  ttctttcaat acatgaattg caaaagatta ataacStaaa tgcaatattg tcttctggat
11581  tggatcttgg aacagaaaaa aggacattag taggaaaatt agtgaaatat aaataaagtg
11641  tttagtcaac aaaactgtac tcatgttagt ttcttagttt tgataaatgt attatggtta
11701  tataagatac tgatattagt ggtagctagg tgaaggatgt acaaaaactc agtattatta
11761  ttgcactttt tggtaaatct gaaattactt caaaattaaa agttaaaaaa aaatgattgc
11821  tgatatccaa tggccacagg tttggaaaac tagaaagaat tcttatccac tcttggtggt
11881  atatatatta gtacagactt ttttttttt ttttttgaga cggagtttca ctcttgtcac
11941  ccaggctgga atgcagtggc gtgatctctg ctcactgcaa cctctgcctc ccaggttcaa
12001  gtgattctcc tgcctcagcc tcccaagtag ctggaattac aggtatgtgc caccacgccc
12061  agctaatttt tgtattttta gtagacacgg ggtttcatca tgttgggcag gctggtctcg
12121  aactcccaac ctcaagtgat ccacccgcct cagcctccca aagtgttggg attacaggcg
12181  tgagccaccg cacccggcag tacagttttt ctagagcaac taggtgttat ttattaaagt
12241  aaatgaatta atacattttg attctgaaaa tctccttcta gaaattagtt ttacataaac
12301  acgaatgcca gattttgaat ttctctttta gatgattcag cataattgct ccaaataaat
12361  ctaaaaggct attgtcaaac agtggaatat ttggatttat aatttcatag ataagactat
12421  cagtccatga tatggttggt gtttaaagta gaatggaagt gaaattccca gagtggagtt
12481  tttaaaataa ccaagttcct aggattttaa ctggaagaat aaatatctga aaggcaatga
12541  ggagtattaa gcatgggatg ccacttctgc accttgagaa aatgaagtat ggaaggatga
12601  gtggttttca tttgagaggg ctgcagggta actgtttcta tttcagatga ggaggtaaaa
12661  agaatgggca ttgaaaagga aaaatgttta ctaagtaggt tatacatttt ggaatagaag
12721  tttctatagc acaatagaaa agtttgagag gtcacatggc aatgatagtt tgggtgagag
12781  tgaggcagga gaatagggtg tggaggcagg gaacctaagg atgtgtcact ccgacttcct
12841  aagaactaaa ttgaaaggaa acccctaact ttccatgcct aagtaacaaa aggaccagag
12901  gctactccct ttgcaaaccc ccacttttc tgcagggcag atgggaaatt ggctgtgggc
12961  agcaaatcag actgattgca ggagagtctt cctttgcata gaagtatcac tttgtaactt
13021  caccatctgc aagaaatcaa actgattgca ggggaaatct tcctttgcat agaggtatac
13081  cactttgtaa cttcacccta gcctctgatt gattgctttt tgcaccaatg tttgcacagg
13141  agcatgacct ttgtaacttc acttcagcct ctggttggct gctttctgca accaatcaga
13201  ctgattgcgc gttacttcat ttacatgagg tgagcattag gtggccaata ggaagcttct
13261  agagggtatt tggactcaag attctgtatc cgggcccttg agccgctgct caggtcgctt
13321  ccgtacagag cgtgctttaa ttttcaataa gtccctgctt ttgttgtttt gttgcttcat
13381  tctttctttg ctttgctggg cgttttgtcc aattctttgt taaaaacgcc aagaacctgg
13441  acaacttgca gtcacgatcc tctaccggtg acaatatcat ttaaatattg gtagtggctt
13501  aaggtagcaa acatcccttt gtggaacgtt tactcagtaa atccttaaaa aacagatcag
13561  aaacaaaaac tggcctccca gggatttgga ctcaggaggt gaattaggcg cggagagggc
13621  ggagcctttg gggtttatgg gcggaattgc acagggtttg gcggggatct ctgctccgcc
13681  ttggcactac tcggctccgc tcgcctccgc gcggctccct gggctttcac tcggctctgt
13741  tcggctcctg ctcgggaagg gtgcgcgggg gcgagtcctg cgcatgcgcc ctcaattcct
13801  gctgctctca gggctgcgct cagcctcgtg ttttcacttc cttttcagag ctgtatccgg
13861  gtcgctgctt tccctattgt ctgaggcagc cgccctcgcg ctgtgcaatt tctggtcttt
13921  cgttgcttct ggtccaggct aataaagttt ttctttcttt aatttttttt cttctagttt
13981  taacgggaga aattaactcc ccggggccgc cgggttgact gcgctgcctg ggccggaggt
14041  cttctccggc cagggagcgc tgtgggaagg ggctcgagcg gccagggcca ggcgaggccg
14101  gggggggcgg gggttagggg accgcggggc tactcttggg agcgcccctg tccggctggc
14161  tgcgcgccgg ttttaaatag catctttcgg acttgtcttc gcggccccag tccccgacct
14221  cggcgctgcc tgggctcctg cagcctctcc ctaagtcttc tccaaacgac cacctcacgg
14281  attccttagt aagtgtatcc gaggcctctg cggcgagagg tccatttcag ccattctaga
14341  agtcagggcc ggtgggagca gggcaggggt gggagagtcc tgcgggaaag caggattggt
14401  ggaccctcgc cctccatggt ccgcgggaat gaagcccgct tgttttattc cctgattttt
14461  gccgttttta ttgactgtta cgatgtctaa gtgggggatt acgaggaagg tgcagaaaca
14521  tctcgaaaag atgagcaaga gggtcagctt cgattttagc tggtctgagt tggcgttggc
14581  agcatcaggg ctggagctgc ctttgaaact gcccgtggag ctgtccagcc ctccccaccc
14641  tgcttctgca ccccaacctt ctcctgaccc cttttgctga gaactcagtt agacaaaaat
14701  tccttgtgcg aggtcatctt tatgttctct tagcaccaca actctcccct gtgtttggtg
14761  tgattatgaa tgacaatgta atttttttct cttcacagtg gatcgcagct ccaagaggag
14821  gcaggtgaag cctttggcag cttctctgct ggaagctctt gattatgata gttcagatga
14881  cagtgatttt aaagttggag atgcctcagg taaatatttc cttctctctt ccccttccc
14941  agctttctgg agttaggctt atttttagac tttttaaatc ggttttgatt aaaatgccaa
15001  ttttaagtga aagcacgtaa acttcattta ttttgtccat ttgatttccc attttgcctt
15061  tagatcctgt cactagacct ataattgctt ctgttttcct cagagataat tagcagtcag
15121  gggaaggact ttttccaaag atgcaaaagg atttctctgt tttacacctt aaaatatatt
15181  tatgcttata ttatagagga aaaaactgaa agatatcccg aagtagaaaa gtatcctgag
```

FIGURE 2-E

```
15241  aggtgtgtgg accattaata gtccctgagc agggttcttc ttttttcaga tcacttgaca
15301  atctcttagg ggagttttgc tttttttgtg tattagctct tttactagaa taaaattgac
15361  cagagtaaga gttgcacttc aaattatagt aggtgctgga ccttgtggat tgggcctttc
15421  agatgtctct gaagtataag tcatgcatgc ttgctcacta tatttagtaa ttttaaaaac
15481  tttttttaaa ggtcgtatat tcacattgta ctaaactgga aaacagaaaa gtatgatttt
15541  cctttgtatt ttctgttcag tgtaccttat gtatatatac agtttcaatt gaaggaatct
15601  gagaaacaaa aattatgttt aatttaaatg ttttatggga aaaatactga taaacatgaa
15661  caatgaaaat tatgtgtaat tccacctctt ttgaaaattt taataaagtg agatgaaaat
15721  ttaaataaaa tgagaagtaa aagcctttcc agtttcatct cttcaacagt ttttaaatat
15781  acttccagag agtttgccgt ggttagtcac atctctcctt cttctggttt tcccttgtct
15841  ttttcctgct agggatagta ggaagggatg aacgaaatta tattactgtt gctacttttа
15901  gtagcatcaa cagcagaatt tgcattgtga tttactgata acttttttgtt tgcataatct
15961  cacttaattt tcacagtaac ttagtgaagg agataccatt ttacaggtaa caggggtatt
16021  gagtttaaag aactggcatg aggtcactca ggcagtaacg gatccaatgg atttgacttc
16081  agaatttagt ctgtttatct gcttggatcc caagagttga tggacggaat cttaaacaga
16141  aactgactat ttggttacta attgaattca tccgcagcaa tcaaaaattg ataagtttat
16201  cttgattaac tgtttttttа tcctttgctt ctcagctctt tatctcccat ttagttggtt
16261  tgctgatatt tggtatttcc aaagaaggga aggggaaggg aaaggaggta aaatttaaat
16321  cttagttctc ttggtaaaga ccttggcaga taagaatatt cctggctagg atgtagtttt
16381  ggtttgttat ggttgtggtt gtaaactttg acaaacatag ttgggtcgtg gaagttacga
16441  attctttgaa tatgggaaca attctaaaac ttacattaag tattacatta ttatgtgaca
16501  ataaatcttg actttatgga caatttgttt ccaaggtttg ttcattgaga tggaatattc
16561  acaggtatca cttctttttc aagtggtaaa acaatctgat acaaacataa agtactttct
16621  caaaatattt tatgatatcg agctaagtag agatttctga ccttgttaaa tcctaattat
16681  agttgaagag aactgttatt tgtgaaaaat gataggatga gttttgttag gttgatatat
16741  ctatatatcc cttaaacaca ctaaaaatat ttactttctg ttccctcttg taatataata
16801  tctagtatgc tgcactcata atttaccttc ctggcccccт ggggagctta aatttgYgat
16861  ctgtggtctc aggtcacaaa tttgtatgta tagttcttgg tatttattgt aaaagggaat
16921  ttagaagaaa atgattgtat ttaaaatgat Yagtagtcaa cagaaattga atcatatttt
16981  tgactcttgt tttaggtgca ttgatgctct gcatagctga gatattSgct tactctagat
17041  taccattgtt ttccatttga atcttttctg tgcctgagat agtatatatt tagttggagt
17101  cttgtagaga ataagacatt agtcctatca ctggtttcca aacatgttga agttgtggat
17161  tcccagccct acttacRaat aggagattaa attggaagta gagaaatggt agctaaacat
17221  ctgtctgatc atatcttctt tcaaaacaac ttctagaaat gactcattga atgaactacg
17281  gacttcctta gaacttaata ttaaagtggt tagattcgtt ggccttaatt ttggctaact
17341  ggattccgtg gatcaatttc ttcttacctt catcttgaaa tctgaaattc tgactataaa
17401  acttttttata tttctgtttg gttttaagaa taaatataga aaacatttgc agataaacat
17461  aatttaactc tatcataaaa tagaaacaaa tcaacattgg taaagaatga cccacccata
17521  agattggcaa gtgattacga gcctgtacct ttagaaaatt atgaaactga agatagttta
17581  tattttattt ttaatgcaaa agcaaaagta agctKctcat ttttgattga aagcagtcaa
17641  taaagtttga aatgagtaag tcctaaaata ggatatatat gcattattag aattatggtg
17701  ataaccacca gaagaaccat atgtagaaat gttaggttga actatatgga attaccagtt
17761  ttgtaggtcc aaaacacaca gtcaaatatt agcaatttca tggttttacc tagttaaagt
17821  gttgatattt gagacttggg catagggttg ggtgtatgat aggagactta tttacatttg
17881  gttatctgct acattaattg aatgttttaa agaatgttta catcttttta aagatgcaga
17941  catgtattaS ttctattaac cagacagatt agccatgctc tctaccccta tcccccacca
18001  ggcaaagtaa aatggggtta actttagatc ttgatcaaaa gttagtttag ttaggccata
18061  ttgccaggaa ataatttaat gagaagtgtc agcctgagac tttggtggta ttttgtgctt
18121  aatctgactt tgaataagtg gaccaagatg tgttcattga ttgtccgtag tttagccctt
18181  tgctagatgt tatggtgaat ttacaaaaca gtagtaataa tattagaata agtaagtaaa
18241  ataacatgta agtaaaataa aaattagaat aattaaataa aatagttgat acaaaggatg
18301  gttgatcact tgtgtagttc aagtctttMa taggaaatga gattagattg atgtagaaca
18361  atttcattga ggagaggagg taaRagtgga actatgaaca gaaggattta gataaacagc
18421  aatcttggtg aggcaaaaat taaaagtgtt gagacttgga attgtttttt tatatttggg
18481  gatcagtgtt gagtaaaagt agaagagata agcctggaaa aatagaataa ggtcaaattg
18541  tggctgtata tacgagttaa Yggtttctca accttggcRc tattggcatt ttgtgttgga
18601  taatttttg ttgtggaggc ttgtcctgta cccagtagga tgtttagcaa caacaacctt
18661  ttcctctact agatgccagt agcatctttc ctacctgccc cagcctcacc tccacctcca
18721  cttccactgt tgtggcaact attgttteca gacattgcca gttgtcctct gggggaaaaa
18781  aatgttcccg tttgagtacc actggctcat acaaaccaag atgacatagc tggaactaga
18841  atcggaagtg gaggatgagt tgtgttgact Sattggctgt cttcaagaca tatacggttt
18901  tggtcaaaca tagccctggg atagctgtaa ataactagta gaatagtgat tgatctgtta
18961  tctattgtgt atatttatag tactaacact gttgtaccag atcaatgaac agaattagat
19021  tgatgaaatt gatgaacaaa tttatattgc atatattgta tattagcata aaactaaatt
```

FIGURE 2-F

```
19081  aaatcatatt gattttgtta tggggagaac attggatata atggtattga tggttgatat
19141  taattaaatt aatttttatt ttattatttg atattgtaaa gggattcttg aaggttggtg
19201  ttttaaattc taattttaaa atttatttta ttttttgtc tggccattac cactcagagg
19261  agtttttaaa aattctgtaa tggtaaaaaa tttcttaact gtctgttagt tgaaatagct
19321  gttgcttgga gagattctga tatattgtat gtttgagaaa ggttatcttt tatttaacca
19381  aatgggaagt aggatttcaa ttttaaagat atatttttt caaacgtaag aaaggtttat
19441  tgtagacttg aaattaactc tttctgccta agataatttc tccagtatat ttctttttt
19501  cctttccttt ctctctcttt ttttttttt tttggaacat agcacagagt cattctttga
19561  tgactaggaa attttgtctt tgcagcctat ggaaaaaata gccaagggcc ttgatttttc
19621  tcattgtcat tattacccag ctatggttgt aaataaatat ggtttatctc cattcctcca
19681  gtgcatatgt acaaaaaaga aatagtgact aattagatga agaagttatt tttcagatat
19741  cagagaaaga taagatttga tgtattgctg atccctatag aaagataaaa tttgatatat
19801  tgctgatccc tgtagttgga gtctgactaa tgtactatat ttggagaatg aggggttggg
19861  taagctatta ggagttggtt ttgtggaaaa atgtccattc ttctaataat agttaacaaa
19921  cataaaacat taaaaatttt ttaaaaaatt gctttctatt caggcatggt tataaatcaa
19981  aaacagaatc aaggagtctc aatttacctc tcatttgaaa aaataattaa ttaattggca
20041  ttggcaacta ccaaaacaaa atcacaaaaa tctctgacat tttgtaaaat ttaccaagta
20101  atagcaaagt tggattggtc tcaattttga tcagctaccc acacttctgc cccagttaat
20161  cttacttttg cgtctatcgt gaagttttga ttgataaaaa gccttctaga aggttgatcc
20221  agaagaaaag tacctttttct ttcatcttca tctcctagct ctcttaatta tggcattgct
20281  tcccttctag agcctaaggt gtttcagttt tctttagtca gtgtgcataa aaaatctttc
20341  tgtgtcttag aagtgttgag tcaactctgg ttattattta ttaaatcata attttaaaat
20401  gtcaagagat tactattatt gttatttta ccagatgtga tttttcttgg taggttgggt
20461  ttaattgctg tgagtggttt acaaaatcat aattttctta atgctttaga gactgaaaag
20521  aagtcacttg tttgtgagta atattgggtc cacactcttg aaatccaaga agcttaaaaa
20581  tccaagtgtt ttgtaaggtt cacacaaact tacttgatgg caaagcctaa cggaactggt
20641  aggagtttat ttgtagtatt tatttctcat actctgtaaa taggatttat acttttctcc
20701  gcagaaataa tgtttgatta taggataata ccatggactt cactgggggt attagtgtaa
20761  tatatggtgc atgttccatg ttacccttga aaatctgaga aattatgaat tctgaaacac
20821  atctggttct taaagtttca gatagggaat tgctgacccg tatgatattg ctcttgtatc
20881  ctatgcaagg ttgtttagga taagagtgtt tttttcagaa cagtcccaag taagctatgg
20941  atacctacat gacagttgtg agattgataa tgatttcata gggtcagtca gtatctgtta
21001  aatgttagtg tttatttttt atgcaatttt gcttgcagga tgtttttgtt agcagtcatt
21061  tgtatgtgac ttgtggggtt cactagatac cagatactaa ttacagacag cttccttcaa
21121  ttttagtgct tagttaatta cagaggttaa gcctagttga gttaatagaa attaaaatga
21181  tgtggaaaca ctttacttttt tccaccattt tatcattacc tttttcttgg ctataggtgc
21241  aaaaaaattc atcatcctgt ggattttact ggctaagact taactggttt cttccccatc
21301  tcctactgct ccctgtcctc tttccctcct cccatttatt tctcgcttca aagattttag
21361  attgttattt acttaattgg gtacaaaaag aatgattact gagtacttac tgagtaatga
21421  ttactgagtt actaagtagt gttgggttga gagggactgg gctgccctag gcaggagtgg
21481  aagaaacctg ctagaaagct tacagtcctg tgagctcttt aatgtcttca gcctaacttt
21541  ttagtaacct ctgtgactgc tcattttcag gatggactgc tttattcatc aatttcttcc
21601  attctagagc atatagtagt tactgcgtag aagttttgaa agtgtagtga ctcttctttt
21661  attgatacga acagtatcct aataaacagg aagattgata atttctcttg caacttcctt
21721  tttctgtgcc attgctgttg tctcttttca cctgtgttct tggtattcct aaaatttgtt
21781  tggtatatta gagttgtagt ataagttgac tatcataaat atctatgctt attttactgg
21841  gtaaagagta aaataaatga agcagtttta tgggttagat ttatcttgtt ttggttttga
21901  ctcaatactt acaagtccat atagtttata aaagatgttt aagggaggaa gaattttgtc
21961  tgtattatca gtaaaattaa aatcaagtgc ccaatcttaa aagaagcaca tttactttaa
22021  aaaataattg ccttttcagt accaaatatt gccatatgac acaaattagt gcttctcttt
22081  ttaaagcata ttttaattta gggttatcta tattcatctt cattagcact atactgaact
22141  aaaaccattg tatcaacttc attgatttat tatttgatca ggttgggaat gtctaccatt
22201  ctttgactta aatttgttta tattgttttg caagttattc acaaattttg gtggtttcat
22261  ttgtgtcagc gttgtgtgtg tgtgtgtgta tgtatataca tatgtgtggt ttttttttc
22321  tcatattttc aacagattct gaagggagtg gtaatggaag tgaagatgct tcaaaggaca
22381  gtggagaagg ttcctgtagt gattctgaag aaaatatttt agaagaagaa ctgaatgaag
22441  atattaaagt aaaagaagaa caacttaaaa attctgcaga ggaagaagta ctatcatcag
22501  aaaaacaatt aattaaaatg gaaaagaagg aagaagaaga aaatggagaa agacctagaa
22561  agaaaaRgga gaaagagaag gaaaaagaaa aggaaaagga gaaagagaag gaaagagaga
22621  aggaaaaaga aaaagcaaca gtatctgaga atgtggctgc ttctgctgct gccaccacac
22681  cagccacaag tcctcctgct gttaacacat ccccttctgt tcccactacg acaaccgcta
22741  cagaggaaca agtcagcgag ccaaaaaaat ggaaccttcg acgaaaccga ccacttctgg
22801  attttgtgtc catggaagag ctgaatgaca tggatgacta tgacagtgag gatgacaatg
22861  attggcgacc tactgtagta aagagaaaag ggagatctgc gtctcagaaa gagggaagtg
```

FIGURE 2-G

```
22921  atggagacaa tgaggatgat gaagatgagg gaagcgggag tgatgaagac gagaatgatg
22981  aaggcaatga tgaagatcat agtagccctg ccagtgaagg gggttgcaag aagaagaaga
23041  gtaaagttct tagcagaaac agtgctgatg atgaggaact gaccaatgat agcctgaccc
23101  tatctcaaag caagagtaat gaggtagatc aacccaattt ttatatctgt ctgtctgggg
23161  aaaagggaat tcttctctaa atcactctac acattgtatt aagtggcttc cttgaaatcc
23221  tatttataga cggctgtggg atgaatgagc accctaactg taacattttt tcatttggcc
23281  aacagacttt tattaaatat ctactatttg tcaccgttct caaggagctc acaatctctg
23341  tcctcaagga gaaatattta gagaaatatt tttatatcca tagacttggg aaatgacaaa
23401  gttttccatt ctccttttta tccctttact cctttgttca caggactgct gaatggtgct
23461  gagtgcatta tgaaaacaaa ggtcaccttt ctttcaactt agttgtgtcc atatccatat
23521  taaaatatag gcctttagaa taccttagag cagggcaaca gtctaatgat tttctgaaac
23581  ctttagagat tcgttttaaa tattaattcc tttaccagtt gtagcactca tcatcttttt
23641  agtcaccaaa tctttgcttt atgtgtatgg gacatacaca taagtatgaa ctatttttac
23701  atatatttag taaataaatt ctctctacaa agaaatggaa taaaaaccca aacatttcca
23761  tcataaaaca gacatttttg tcacaaaact gcagtaatgg ctcagtgctt ctacttgaag
23821  ttctactact gaaagttagg tttattcatt gatctcagag atgcatcatt caagactctt
23881  cagtcaaaca catttgtgtc tctttgtctg tgtgtttttg tatatgtttt tcttcaaaag
23941  aaagattgag ttgttgtctt acatgtagat gtatctgtta catcacatgt agaactgtaa
24001  taattctgtg atttctgtta atgtgtttat attcaccaaa atttatttct ttaccaaaca
24061  cttatttttt tccatatcac tgaagtaagt taaacgggta gaaagatttt tcctcttatg
24121  ctttatttgt tctttcacat ttttctgcct acaagtttta catagcagtg attatctatt
24181  ttaattgctt tgcttaaaaa tttggtgatt tcgttaccaa aataatctat agcaggctgt
24241  gaggttatca gtggagatga gacagcagtt tatatttgga ttataaacat tggttaactt
24301  ctgaaaccaa tattttatat gaaaaaatat tgcctgcctt caatgtcatt ctaattctta
24361  cctttaaagg aaagttgtca tttgggaaat agtgtaatat ttttaatgat atagcacagg
24421  cagttccata ttatcaatgc tatcttaaac aggcatggaa aaatctaatt gtttgtattt
24481  aattgaggca tgcatgcagt ttacgttata gcagttaata ctctatatat aatgaataag
24541  tttacttttg accacatcac tttcttcatt tttattaggt tgatattctt agaagtagtc
24601  tagctaatga tagaggcagt gtcgattaga tttatgctaa aatgtggggg acacctgcat
24661  gtgtttagga atttgccacg taaagtgtgt gtagtatctg tttattaaac ctgtacttca
24721  taaatactgc ctaagaggtt gtagactaac atttatgatg catttaaagt ccatgtgttg
24781  gttgtttagt tttttttccc cactagaaga cactttagta aataaaggga ccttttatgt
24841  ttaggttaat agtcccaaac acaaacacat tccaggaatg agattccaaa gtgcttcttc
24901  gaaaagtgcc catgtagaag agattataaa aaggtacttt ttctccagaa taattctcct
24961  taaaagcttg agtcattttc ccctcaattc tattattcaa gctctacagt agtttttga
25021  tgattttagt ttatttaaat tatgtccctc aaattagata attacatttt aattaaaaaa
25081  ctaatcaatt gtgattatta gttatatttc tagttttctt atttgtatga tcttaatttt
25141  gtgaaaagat gattaaagct agtccttgtt gcctgtggcc aatgtttggg tcttaatctt
25201  cctggttatt ttagctagtg ataatttctg tgtccaatac atagttaagt atatatttga
25261  ttatatttga agtttgattt tccctgcttt tactttgttt tcagaaacta ctttgtagta
25321  gagttaggtg aaaatgtagc tttagaaata cataatagaa atatataatt ataccatatt
25381  tagtttataa gcaggcagac atttttctta gctgctaagt tcatttgtag caaaataatg
25441  acctgaaact ttcataattt tgaagtttct taagggaact tgatatttat tttagcaggg
25501  ccctttata attactgctc ttaatggcct taacatatgc ttaatggcct taacaacatg
25561  ctgcagaccc tgaaatactc tctagcagtg attctcaaac tgttttttt ttttttttt
25621  ttttttttt tttggagaca gggtctcact ctgttgccta ggctggagta cagtggcatg
25681  atctcggctc actgcaacct ccgcctcatg ggttcaagca attcttctgc ctcagtctcc
25741  caagtagctg ggactacagg tgcatgccac catgcccagc acatttttgt atttttagta
25801  gaggtggggt ttcaccatgt tggccaggct gttctccaac tcctgacctc aagtgatcca
25861  cctgccttgg cctcccaaag tgctgggatt acaggcatga gccaccacac ctggtctcat
25921  tttaaaattt tttaaatgtt tatttttttt gagacagatt ctcgctttgt ttcctggact
25981  ggggtgcagt ggtgtgatca tggctaactg cagctttgat ctcccaggct caggcaattc
26041  tcccacctca gcctttggag tagctgagac tacaggtgta cgacaacgac tggctaatgc
26101  ttttatttt gatttttttt ttttttttgt agagatggta tatcactatg ttgcccaggt
26161  tggtgtcaaa ttcttgacca caagcagtcc tcctgcttct gccttccaaa gtgttgggat
26221  tacaggtgtg agccaccaca catagccccc ttacagtttt aaaaattatt gaggacaacc
26281  attaatattt gtgtagatta tatctgttga tatttaccac ttcaaattaa aactgaggaa
26341  gcacacaagt tatcagatcc catttcatgt agtgtctgga aaacttcatg gtactcccat
26401  gagagaaaga gtgaaaaaga gaaataatgt cttagtatca ttatgaaaac tgcctttacc
26461  ttttatacct actgaaagag agttcactgg accacacctt gagaactgct gctctagaac
26521  tgggcattag ataaactgca ttttccgcaa acggacacct gtttttgtca ataaccttat
26581  ggcatttttt aaatttagtg cagcctgaag aatccccttg aaatgtggtt gtatgtcttt
26641  ttcagccgtt aatcttgcag ctttgagatt ccttaagatt atcttgccct ttagttttct
26701  gaggtaagtt ttttggggac ggcagtgtac tattttcaga aatggttccc attttcaaga
```

FIGURE 2-H

```
26761  ttttgaaact aaagttttgc atgagaatta atccatgacc gtaaattata tcaatagaat
26821  tggcattttt gttttgtaaa agaaagaaga gctctgttct cacagaaaat atttaaaatt
26881  agagacagtt tggattttgg attactgctt tataaatgta actttatgtt ctttgctgag
26941  tagaattttt tttccataaa agtctccgtt aagtacatta gatgtcatgt tgagaaatgg
27001  caattgtcac ctgatttttg tgaattaaaa tgtaaatgtt ttgccaaaat ttataaatat
27061  ccacttctat ttgggacttc taacttatta atcagttgtc tttttacaca ttgcaaaatg
27121  aaacaaattg ctctgagtcc taaaaaaagg tYagttctgt aacatttata aatatcgaga
27181  ttcatgagaa tttaaccttt tctaaaaatg tagaggaagt atgattggtc aagctttttg
27241  ttataataat tttttcttat cttcagggca gtaggattct tttaattcta attttacagg
27301  aaatgYgctg gaatttcctt gtattaaaca catgtgattt agataaaatt tagcctattg
27361  tatatatcat ttacagtcat atgtggaata tatatatata aatgaacagt aagtatattt
27421  tagagcctgt ggagaacatt gatcagtatt atcattttgg gggtaagttt taaatcctga
27481  tggaaagaat gcaataattt actataagta gtgattgatt tattcatata ctgtcattta
27541  ttcttcctat aattacattg gattccatta aattattcca tcataataaa gttatttttct
27601  aggtggtact tttttgccat tcaaaaccta ttttcctgat gtctattttt atgtaaaacc
27661  atattttaga aattttatgt taaaagcttc agcttaacta ctttctctga aaccttgaaa
27721  gatgatatac gtcttcagaa atatactaac aacagttgat aataaatcag gttttgtatt
27781  gatctaatca taagtgtgct taaacttgtg tttgatctgt taaaacagat ctagctgggc
27841  gtggtggtac atacctgtaa tcccagctac tcgggaggct gaggcaggag aatcgcttga
27901  acccgggagg tggaggttgc ggtgagccga gatcccacca ttgcactcca gcctgggcaa
27961  caagagcgaa actccttctc aaaaaaaaaa aaaaaagaaa aagaaaaaaa agaaaactga
28021  gatcctaatt ttttttaagg aggatgtgtt acttagattt tccagttgaa acattgtctg
28081  ctagccattt Rgttagggaa atattttatc tctagttttc cctatttccc tctttgcgtt
28141  acatttctat taagagcctc aagtcatgag gatcagaggg ccagaaacat tagttaatgt
28201  tgttttctct gcctctaaat gataatagat gtggagagaa attggcgtag agatatgatt
28261  taggcattta tggtttttct gagcaaaatg gaaaaatagg aaaatagagg aattagaact
28321  gcattttgat acaacttttt ctgaccttta gaaatttact tgtttcccat atgggtaaac
28381  ctttttttcca gggctgaata attttaattc caccagagaa attaccacac agtccataaa
28441  gacatgtagc atcctaattt tcaaactgtg attgtttgct aaattagttt ctagtctgtt
28501  ttttttttttt tttttttaa cctagatgtg gtattccaag gaaagtcttt tttggactta
28561  ccgtggggat gatatttaaa gtagtagtgc tcattggagc ttgcaacttt tcttttgggg
28621  gtacaaggaa gatcctcagt gtaataaatt atcagttcat Kttcctcttt ttgaaccacc
28681  ttaacaagag atacagcgga aaggaaaagg cattgaccat ttattgagtg tctagtttgt
28741  gccaggaaca ttcttgccat atattgttaa ttcttataac ctcctgaagt agatattatt
28801  tttatcgcca ttttgcagat aaagaaatta cagcttcgag gttcagtagg gaataagcga
28861  gaggtgagaa agaaacttag gaatgtttga tttcaagtca actttgaagc atggctagta
28921  gtatgaaaat ctgggcagat ttttgtttgc ttactctggt actgtaaaac cagatgagtt
28981  ttctttagga tgttttcagt tttatttgcc Rgccatttgt cagccccaga cctattttat
29041  ttggttttat aacacagaga agcaaaatac agaacttaga aatactagaa ttctaaagag
29101  aaatgatact atattcttgt tgatgtttgc attcttttct gattgtttta caaatctgcc
29161  gcaattttag atgttataca caaagtatat tggaagaatt attaattgtg accttcaagt
29221  gtattcaagt gtcagttaca aattgctgtt gaaatgatgg ctccaaatgt tacttttgga
29281  tattggttga gtttaaaaga atataaaaat ataaaaaata ataaaacata tttatgaaaa
29341  atatatattg tgtattgtat gaaactaatt tattaaattt atttctgaat attgttaggt
29401  tcataaataa aaatgtatgg acccaagtgc catcttttat tatatgtaat gttaagcaca
29461  atctattaaa tatgcaagaa tcatttttctt aacttactta aagtttaaat ctataatgca
29521  gaaggtttaa atgcaaacaa gcagaccctc ttctgaaatt ttatctttta aatattagtt
29581  tattaatacg tcttcctata ctttttcttt ttgagacagg gtcttgtgct gtcccccagg
29641  ctgaagtgca gtggcgagat aatagcttgc tgcagcattg aattcctggg ctcaagtgat
29701  cttcccgcct tagccacctg agtacctggg actaagggca tgcatcacca cgcccagcta
29761  attttaaaat attttgtaga gacaggatct cactgtgttg cccaggctgg tctcaaactc
29821  ctggcctcaa gtgatcctcc tgcttcaacc tcccactgta tttgggatta ccggtgtgag
29881  ccactgtgcc catcctctct ttttatagta tcactagcta gatgtattta ttaccagttt
29941  acactggata tttcttgata gMgacctaMt taatagagtt ctctgaccct tctagcagct
30001  agtcaatctg gatccttttt tagattattt tctgtatctc tgtgtattca tatacctacg
30061  caggctctca tggatctgtc ttacRgtagt ttgataatgc tggaatttct gaatttcccg
30121  gtacttaagg agtatagagt ccttaagtgt gccatcacat acatacttgc ttatatgtta
30181  ttttgaattg attttttcct attttgtata tgtgcgtctt gtcttctcta ggtaggatta
30241  taagctccta aagtgtaaga atcatgtctt gtgcttttaa aacacatgta taggatcaat
30301  gattactttg attacttgtg gattcatttt tcgtttttcc tagtttaagg ggcaaagcag
30361  aatagcaaag cagatcaaca gtagttagtt gcatgcgtgc ttttcatat cctagaggag
30421  agagaatacg aagcttgcag agaagtgaaa aattaaataa gagggtgctt acagggttct
30481  ttgataaaga ggtcaagtat atatttctca aatataatgt actttgaagt taataaaaga
30541  atggaaaatg gggcatacat ttggaaatgc atataataat gtttaagtaa taaagaattt
```

FIGURE 2-I

30601 tagtgattaa gtttttttt tactttaaat gtgattttct gatatatgtt aatattttaa
30661 tattaggact cgctgattct tgagaagagt caaaactgga gctctcaaaa aatggaccat
30721 attctgattt gctgtgtttg tctgggagat aatagtgagg acgctgatga aataattcag
30781 tgtgacaatt gtggcattac agtccatgaa ggtaatgttg ctttcttttc tctcttttta
30841 gaaatggctg actgtggctc ttttacattt gtactaaggt ggttctacaa tatttcatgg
30901 tgtttcggtt gaaataatgc taaatcatca aagtatggat gctatttttc aggttatctt
30961 ttcttttatt ttgagatgga gtctcactgt gtcgcccagg ctggagtgca gtggcacaat
31021 ctcggttcac tgcaacctcc acctcccagg ttcaagcgat tctgctgcct cagcctcctg
31081 ggtagctggg actacaggcg cccaccacca cacctaattt ttgtgttctt tttttatttta
31141 tttttattta tttattattt ttgagatgat gtctcgcact gttgcccagg ctggagtgca
31201 gtggtgctat cttagctcac ttcatcctcc gcctcctggg ttcaaactat tctgcctcag
31261 cctcttaagt agctgggatt acaggcacga gccaccatgc ctggctaatt tttgtatttt
31321 tagtagagat ggagtttcac catattggtc aggctggtct tgaactcctg acctcgtgat
31381 ccacccgcct tggccaccca aggtgttgga ttacaggcgt gaactaccgc gcccggccaa
31441 tttttgtgtt cttagtagag atggggtttc accatattgg ccaggctggt ctcaaactcc
31501 tgactttaag tgatccgcct gcctcagcct cccaaagtgc tgggattaca ggcgtaagcc
31561 accgcacctg gtcccccagt catcttaatt aaaaaaaaaa ttttaaataa tgaaaataga
31621 aatgctatat gaaaatcagt gtttctttaa aggtcRgtgt tctggtattg tattttatgg
31681 tactRtttct tgtttgtttt tttttaattt tttaattttt taatttttt ttttttttt
31741 tggtagagRc agggttttgc tatgttgccc aggctgatct tgaactcctg gcctcaagca
31801 atctgcctgc ctcggcctct caaagtgctg agattacagg catgaaccag tgtgccaggc
31861 cttatttcca gttttttttt tttaatttta aatttttaaa aattatttct agtttttaaa
31921 taaaattcag aagtaggaaa aagttgacta ttttagtctg atagaaaaca ttcttaaaaa
31981 atttatattc tatcaatttt ttgtagcttc ttctttcata gttgtatttt tatttttgag
32041 acagagtctt gctcttgtcc aggctggagt gcagcgacac aatctcaggt cactgcaacc
32101 tctgcctccc aggttcaagc gattctcctg cctcagcctc ccgagtagct gggattacag
32161 gtgcctgcca ccatgcctgg ctaatttttg tatttttagt agagatgggg tttcaccatg
32221 ttggccaggc tggtctcgat ttcctggcct caagtgatcc acccgactca gcttcccaaa
32281 gtgctgggat tacaggcatg agccaccgtg cctggctgac agtatgtatt tgttaagagc
32341 aaagggccgg gtgcagtggc tcacgcctgt aatcccagca ctttgggagg ccgaggcggg
32401 cggatcacct gaggtcagga gtttgagacc agcctggcca acatggcgaa accccatctc
32461 tactaaaaat acaaaaatta gccgggcgtg gtggcttata cctgtaatcc cagttactcg
32521 agaggctgag gcaggagaat cactggaacc caggaggtgg aggttgcagt gagccgagat
32581 ggtgccactg tactccagcc tgggcgacag aggaagactg tctcaaaatt taaacaaaca
32641 aacaaaaaaa ggcattctct tacataacca cataattatc aaagtcagga agttagcatt
32701 gatgaaatgc tgttttatct aatttatata gattgtattc agatgttgcc aattgtccca
32761 acaatgtcat ttctaacaaa atcaaagcca ggatcatgta ttgcatttaa tcgtcatgtc
32821 tgtttagtca cctttaatgt agaacagtcc ttgagtttct gtctcataga cattttaaa
32881 tatatagtcc agtttttttt attttggttt gtttttacag aatgttccta atatgattag
32941 atttgggtta tatgcttttg gccggaatac cacggaagtt atatagtgtt cttagtgcat
33001 cgtattagga agcatatgat atttatttgt gcagttactg gtgatattac tattactttt
33061 tttctcattt ttattgcata agcctaccaa tttagaatac tggtgatatt ttgatcactt
33121 ggttgagagg gtgtctttct ggtttatctg tggtaaagtt atttttttctt accaattaat
33181 tattattatt attattattt tttaatcagg gctgtgacta ccacaaaaca gtatttctat
33241 ggcctttctc ttacctgctt ttctggtatt acagcccttg tcccaagtaa ttctgggagt
33301 ctgctggact tgtgggactt tgcatgatct gtgattctct agctaatgct ctgatcttca
33361 tgtggagctg ccctttttttt gctgctaaca ctcactgtgg gaatatctcc agttactgtt
33421 attcatttat atttttagta atgtttgttg agtaccttct atgtgcttgg tattctgtaa
33481 attttaggca tatgaaatag aatgtacaca ttgttgacct ggtggaactt aatgtctgga
33541 cagacaggca gtaaacaaac caataaatat ctcttagttg gtggcccaga ttctgaagaa
33601 aaatgaagta ggacaaggag tataggaatg ctgtgtgtga gggtggttgg gtgagttgca
33661 gaagtttgta tttgtaaaaa ttgataaggg gctgggtacg gtggctcatg cctgtaaatc
33721 ccagcacttt gggaggccaa ggcaagcgga tcacctgagg tcaggagttt gatatcagcc
33781 tggccaaaat ggtgaaacga cgtctctgct aaaagtacaa aaattaatcc ggcatggtgg
33841 cacgtgcctg taatcccagc tactctggag gctaaggcag gagatttgct ggaacctagg
33901 aggtggaggc tgcagtgagc caagattgcg ccagtgcact ttggcctggg tgacagaggg
33961 tgactctgtc tcaaaaaaaa aaaaaaaaaa gttgttaagg gaaatccact ctgataagag
34021 gatatttgag cagtgacctg aaggaagtaa aggagcaagt cacttaaata tattgggaac
34081 gagcactgca agtggaggga gtagtgaatg taaagatctc agccaagcac aatggcccac
34141 gcctgtaatc ccagcacttt gggaggccga ggcagtgaat cacttgaggt cagaagttca
34201 aggccagcct gtccaacatg gtgaaacccc atctctacta aaaatacaaa aaattagctg
34261 ggcatggtgg catgtgcctg tagtcctagt tactcaggag tctgaggcag gagaatggct
34321 tgaacccagg aggcagaggt tgcagtgagc tgaggtcgag ccactgcact ccagtctggg
34381 caacagagtg agactccgtc tcaaaaaaaa agatcttgaa gtggtagcat gcttaagagt

FIGURE 2-J

34441 gtttgtataa tggcacagaa gtttgtatgg ctagaactgg gtgagcaagc tggttagtgg
34501 tgggagtaga gatgggagta atccagctgg ggtcttgcag attccggtac tgattttaga
34561 ttttattatt gtgtccggaa ttggtgggtt cttggtttca ctaacttcaa gaatgaagct
34621 gtggatcctc gcggtgagtg ttaacagttt ttaaaggcgg cgtgtccgga gtttgctcct
34681 tctgatgttc ggatgtgttc ggagtttctt ctttctggtg ggttcgtggt ctcgctggct
34741 caggagtgaa gctgcagacc ttcgcggtga gtgttacagc tcataaaggc agcgcgtacc
34801 cgaagagtga gcagcagcaa gatttattgc aaagagcgaa agaacaaagc ttccacagtg
34861 tggaagggga cccgagcggg ttgccattgc tggctggggc agcttgctct tatcccctc
34921 tctggcccca cccacatcct gctgataggt ccattttaca gagagctgat tggtctgttt
34981 tgacagggtg ctgattggtg tgtttacaat ccctgagata gacacaaaag ttctccaagt
35041 ccgcacagag cactgattgg tgcatttaca aaccttgagc tagacacagg gtgcagattg
35101 gtgtgttaca aaccttgagc tagacacaga gtgctgattg gtgtatttac aatcccttag
35161 ctagacataa aggttctcca agtccccacc agattagcta gatacagagt gctgattggt
35221 atatttacaa tcccttagct agacataaag gttctccaag tccccaccag attagctaga
35281 tacagagtgc tgattggtgc atttacaagc cttgcgctag acacagagtg ctgattggtg
35341 tatttacaaa ccttgagcta gacacagagt gctgattggt gtatttacaa tcccttagct
35401 agacataaag gttctccaag tccccactag actcgggagc ctagctggct tcacccagtg
35461 gatcccgcac cagggccaca tgtggagctg cccgccagtc ccatgccgtg tgcccgcatt
35521 cctcagccct tgggtggtcg atgggaccgc gccacagagc gggggcagca ctcataaggg
35581 aggctcaggc cacgcaggag cccatggcgg ggtgggtaag gctcaggcat ggtgggctgc
35641 aggttccgag ccctgccccg cagggaggca gctgaggccc ggcgagaatt tgagagcagc
35701 gccggcgggc cagcactgct gggggacccg gtgcaccctc cacagctgct ggcccaggtg
35761 ctaagccccct cactgcccgg ggctggcagt gcccgccaag cccatgccca cccggaactc
35821 gcgctggccc gcaagcgcct tgcacagccc cggttcccac ccgtgcctct ccctccacac
35881 ctccccgcaa gctgagggag ctggctccac cctcggccag cccagagagg ggctcccaca
35941 gtgcagctgc aggctgaagg ctcctcaagt gcggccagag tgggcgccaa ggccgaggag
36001 gtgccaagag tgagcgaggg ctgtgagggc tgccagcacg ctgtcaccta tcattatgat
36061 ggatgtggga agccataaca tggcttttca gcagaggatt aatatggttt gattcatgtt
36121 ttggaaggat cactcagact gctgggttgg gagtagactg aagtgaatct aaggtgggaa
36181 actaggaaaa tatttaggag accattgcaa ttatctaaga gaggaattgg ttttgaggta
36241 gatagtaaaa gtttggtttt aaactctgtt tatgtaagat gttaggacta gaaatttaaa
36301 tttgggactt atcagcctga agatgttatg caaagatgta agactaaatg cttgtaacaa
36361 agtcatccta gtactttaga ggagatgggg aacacatggc cagtgagata tgaggaccat
36421 gggtgatgac ccagagaaaa ctgtttcaaa aagagagtca gtaaatgtgt caaataatgt
36481 ttttaagagt gggaaatcac cctaattcaa tgctggactt caagtacacc accaggcaag
36541 ctatcccaga ggttattgaa aaaaaaaact aattgataaa atgtttatat cacatgattc
36601 attaaataaa tataaaatac acagaagcaa ggggaaagat agaggataag ttaactcatt
36661 taggctgggg tttaagtagg atagaaggac cacactacct aaataaatcc tggggtacat
36721 agtgttgata tgttttggct tatcaacctt ctttgctaat gctgcggttg tgagaaccag
36781 agcccacaag ataggtgccc agggccaaag ataggcacac tctcaggaac atacaggagc
36841 aagtatgcct ggtcatggcg ttagctctcc agttaatcta ctcaacagcc gtaggtttta
36901 tttttatttg ttagtggaca gagtaggtat caccaatggg tggccaaatt aaggcctcag
36961 gaatattgag ttctaagtgt cgcatatatg tttaatgtat tctgactctt tggtgccttt
37021 tgcatatttg cctcaagaat attagataca tcttggtcct tctatccctt tctgtctgta
37081 tttaatgcat acgtgcttta cttacttttt atctatacat acacatataa gttttattga
37141 tctcattttg tttctctata gtagagttta atatcctcaa gcagagcaaa aagacattat
37201 tactgagaat atatcactga atttagtaac atgagggtca ttggccatct tgacctgaga
37261 gagttcaaag tgagaataaa agaagggaaa gtgcaaaaag tgaccaaatg caattcctgt
37321 gatgagtttt tatatagaga agccaggaga tgggacaagg gatgtgaaat tgagggaggg
37381 tttttttctt tcttatttta gatgctgtat tgattatctg ctgcagtgga acaaattatc
37441 ccaaaacata gcagcttaaa acaacaaaca tttattattt cacatgattt ctaaggggta
37501 agaagctggg agcagttggt tgggttgttc tggcttagag tctctcatga agttgcaggt
37561 caaagccata gcctgggctg tagtcatttt aagatttggc tggacctgga ggatttgctt
37621 ctaagctaag tcaaatagtt gttggtaggc ttcagttctt tgatggctgt tggccagaga
37681 cctcagtttc ctactgcatg gacctccctc ataagctgca tgaatgtctt catgaagggc
37741 agctggcttc agagtaagtg atcaaagaga gaaagagacg gaagcagttg tatcagagtg
37801 ctagtatatt ttataatcta atattggtag tgcatacaac catttctgcc atattccgct
37861 gttcacacaa gccaacccta gtcagatgta ggtgggggca ttctaagagt gtaaatacca
37921 ggaggcactg attattaagg gctatcttag aggctgccta ccacatgtgc agagattata
37981 tagcatgttg tgtgctcatg gaagtgaccc agtggaaagg gagaattggt gatataagga
38041 aaatagtgag gaaagaaggg ggaacaattg ctggaacaaa ggttttgagt gcaagattag
38101 aggggatgag atccagtacc aagtagagag attgagtaga taagagtaac catagtaatt
38161 agggaaaaaa ggatatatgg gtccagatgt agctaggttg ttagatgtgg tagttggagc
38221 atgtgtaagt tatttttaaa atgtttcttg ttttctcagt gaaataggaa gcaaagccat

FIGURE 2-K

```
38281  tagcagagag tgggacagtg gggagttaga gatttgggga ggcagagttg tataaattta
38341  tttgggagaa taaactgaat aaggaaacat atacctaaaa tatatatgta tacacacaca
38401  gacacatata tgtatatata cacacacata cacagcaata agaccacttg atattagtgg
38461  taatgaatat aaagagctca ttaatataaa tatattcata tgaatatgat attagtggta
38521  atgaatactg ttactgagaa aagtaaacat ggttgaatat atgtgtattt ttcttctcca
38581  ggtgtggttc agattcctgg gagtaggtaa agaatatgtt aactcaggag tagcagctac
38641  tttggcagtg tttgttacat tgatcactga tgatcacaaa catgttactg ctgcttatat
38701  ccatgaatgt cattgctgat gtgtgataca gtcatgatga gttctgtcag ttttggtgat
38761  tctcttatga tgctgagggc actatctcta tttttctctg cttcattata atgctgagta
38821  ccatctctac ttcttaagtc attcttaggt ctacttctta gattgtggtc attgactatg
38881  gatgtatttt tttaggccag cacagtagtt ttgagtaatt gaattggttg ctaacattta
38941  gacattggga aatcgtacac atttaggatt tcaagcttct tttgcaaagt tctggcggca
39001  ctagggtttt ttgctgttgt tgttgttgtt aagtgatgga gtgccccagg cttgcctcaa
39061  actcctgggc tcaaacaatc ttcttgtttc agcttcctga gtagctgtga ctaaaggcat
39121  taaaggcata tgcctgccct tgcagtagtt ttcactttct tctagcacaa tctgctggag
39181  ctgaataagg ctgccccctg atagtaaggc ctgtggtctc cagttcactc cagtcctctc
39241  tcaaaccact tactcatttt tgttagttgc ctagtccctg gagatatttg aatatatagt
39301  tccttatctg ctttgagggt actgtctctt gttgctgata tttgggctgc tgatgccttg
39361  tggtctcaat gtctcacatc tataatagta aagcaactct tttactatag cttttacttt
39421  aggacacaag cttgagtcca ctggggacta gttccatgta catttcaaaa tgaccctcgc
39481  tctctatttt gttgtttttg agtcatgttt tgctccttgg actcttaact taggatgagt
39541  tctttcccag tagattcatt gttcagataa acatagattt ctgtaagggc acaacattag
39601  ttccactgac atgcttacag ctattgaaag ttcataatac gacccctgct tcagatgaat
39661  agcctcggta ctggaaaatc taccttcttt ccaaagtagc tttgattttc agaaacattt
39721  aaaaagtctg aatcaaggca gacaaatgag atgatcaagc tgagtggtaa tgccaattaa
39781  gtcaacagga tgttgtgtct atataaagtc atgtaacaac tcttcttgtg taattgattt
39841  aactgacttt gatggcagtc caagttataa atatttttga gactttgcta atgataactc
39901  atttaattcc taaaactaat tcattagaat tggagaccta tttttgttac agagaaggaa
39961  gaaagtagag aaggagaggg aaattgatgc ataaataggg aagctgaata gtctttgacc
40021  actttaggtg aatctttatt ttgtataatt aatttattgg tagtaatttg aattctgtga
40081  actttatata tgttccaagt aaagatgcaa aagatgcaaa aaggtcaaaa atacgttttg
40141  acttaaaaaa ttacgctact actttcatcc tctagtttaa ataatttccc atttccttt
40201  ttaattttgg agtttctgtt gtatgattga ttttgtatgt tgtattatta gactttctgt
40261  tgtatgatgt ggagtatctt ttcatatgct tatttgcctt ttttttttt ttttttttt
40321  tgagatggag tcttgctctg ttgttctggc tggagtgcag tggtgggatc ttggctcact
40381  gcaacctctg tctcccaggt tcctgcgatt ctcgtgcctc agcttcctgt gtagctggga
40441  ttacaggcac acaccactac acctgactaa ttttttattt ttagtagaga tggggttttg
40501  ccatgttggc caggctggtc tcaaactcca gacctcaagt gatccacctg ccttggtctc
40561  ccaaagtgct gggattacag gcgtgagcca ccacacccac ccttatttac cttttgtata
40621  tcttcttgga tgaggtgtct attcagatct tttgcccatt tttaaaaatt ggtttgtttt
40681  cttattgtta agttttaaga gttccttgta tattttggat atcagttctt tatcagatag
40741  gtcttttgca aatactttct cccagtctgt agtttgtctt tttattctct tgaaggtatc
40801  tctcacagag cagaagtttt acattttaat ggagtccaac ttataaatta tttctcttgt
40861  ttattacgct tttggtcttg tatctaaaaa gtcgttacta tgcctgaggt tatctagatt
40921  ttctcccgtg ttatttttcta gttctataat tttttaatta taaatttttct taattaaaaa
40981  tttaaaaatt tttataaatt tgcatttcac atttgggtct gaatttattt ttgttacgag
41041  tatacggtct gtctcggaat tcattatttt gcatgtgtgt gttcaatttt tctatcacca
41101  tttgttaaaa aagctgtctg tgctccctta tattgccttt gctcctttgt caaagatcag
41161  ttgactatat ttctttgggt ctttttctgg gctcagtttg ttcatcaact gacctatata
41221  tctattcttt gattaatacc acactgtctt gattactgca gcttcatagt aagtcttgaa
41281  gttgagtagt gttactcctt tgacttgttt tttttccttc aatattatgt tagctcttct
41341  gggtttttg cctctccatg taaactttag agtcagtttg cagtatctac aaaattacat
41401  cctggaattt tgattgggat tttattgaat ctatagatca agttgggaag aaatgacatc
41461  ttgacaatat tgagcttttt tattcatgaa cataataaac atttatttat ttagttcttt
41521  gatttctttc atcagagttt tgcagtttc ttcatagaga tcttggcaat attattttat
41581  tttttggggg gtgctaatgt aaatggaatt ttgtttttaa tttcaattct acttgttcat
41641  tgctttcata agaaagcaat tggcttttgt atattaagtt tgtatgctgc agtcttgcta
41701  taattgcttt ttagttccag gaaatttttg gctactcttc cagattttct acttagacag
41761  tcatgtcatc tgaaagcaaa gataatttta tttctttctt tccagtgtgt atatttttgt
41821  ttccctttct tgtcttatta cctgatgttg aaagcagtac tgacagagaa catccttgcc
41881  tcatacctca tctttgtggg caagcttcta gtttctcatc gttaagtatg atgttagcta
41941  taggtttttt ttgtagtttc tttatgctct agttgaggaa attactcctt atttctgttt
42001  gtgtttttt tttacttgtt tgttttgaga cagagtctca ctctgctgcc caggctggag
42061  tgcagtggcg caatttcggc tcactgcaac ctctacctcc tgagttcaag tgattctcct
```

FIGURE 2-L

```
42121  gcctcagcct cccctaatgt gtataatttt aaaaggtgcc tcaagattcc tcatgcagga
42181  gatgggaagg atcatccttt gagagacatt ggtttagaag aaagtaaaaa tgggggggata
42241  aaaggaactc tgcaatgtgg acgatgtcat gctggagctt agacagccac aggtgtaggt
42301  caaggggggа gccagagcaa tgtagaatga agtatggttc ctggtataag gcaaatatca
42361  ggttgacagt gaggtttcta aaaattaaga ctttttattt tatttattta tttatttttg
42421  agacggagtt ttgctcttgt tgtccaggct ggagtgcaat ggagcaatct cgactcactg
42481  caacctccac ctcctgggtt caagcaattc tcctgtctca gcctcccaga gtagctggga
42541  ctacaggcat atgcgccacc acgcccagct aatttttttgt atttttatta gagacgggat
42601  ttctccatgt tggtcaggct ggtcttgaac tcccccgcct ctggtgatcc actcgcctct
42661  gcctcccaaa gtgctgggat tacagacgtg agccaccgcg cccggccaaa aattaagact
42721  tcttaatcct cttaataatg gttagggcaa acacaggaaa gaagaggcaa aaatgaaatg
42781  ctgatgacaa aaattaacgg tgtgagcacc aatagaaggt tgagggtcaa tgcagatgga
42841  agcttgggta gtatctgcta aagagatgtg ccacctcttg ctccagctct tctactgtct
42901  tcagggtgat ttttctttct tttttttttg agacggagtc tcgctctgtc accaggctgg
42961  agtgcagtgg catgatctcc gctcactgtc gcctcccagc ttcaaggaat tctcctgcct
43021  cagcctccca agtagctggg attacaggcg tgtgccacca cacccagcta atttctgtat
43081  ttttagtaga gatggggttt caccatgttg gccaggatgg tcttgatctc ctgacctcgt
43141  ggtccgcccg ccttggcctc ccaaagtgct ggtgcatatc tgttcatatt attttctgct
43201  taaaatcagt cccttaaaat aaagttcacg ctaatagcag gactctctga gatcccatac
43261  cctgaccatt tgaatgactt gtctttttct ccagtgtttg ggggctttca tatatgcttt
43321  tcctttttacc cagattcctc ttgttccaca aatataacct aaatacatag gtgttaagcc
43381  agtgtttacc gtttgtcttc ctgttatttg gagatttgaa aaatcttact tttaatttta
43441  tttttaaata ttttattgtg tatatttcag gtatacaata tgttgttatg ggatacatat
43501  agatagtaaa aaggttactg tagtgaagga aatgaacata tccatcatct cacataatta
43561  cccaattttt aaaaagttta ccaagtccat ccttctgcag aaattaccta ttttttgtt
43621  tttgtgacaa gaacagctaa aatttacatt taacatgaat cccatacaca gtacaagttt
43681  attaccttga attaattcat taaaatattc ttaattataa gtgacatatg agtagatatg
43741  taaaaatagt tatgctgtct gtctcaactc tacttctctt tgcataggga atcactgtta
43801  acaatttggt atgtatctac cccatctgtt gtaccttgtg taccaggggt tttttaactt
43861  tattttatta tgtagatata tgtcctctgt ggaaaacatc agtagaaata agctaagaaa
43921  acatctccca gttgggcttg gtggctcacg cttgtaatcc cagcactctg ggaggttgag
43981  gtgggtggat cgcttgagct caggagtttg agaccagcct gggcaatgtg gtgaaaccct
44041  gtctctacaa aatacagaaa aatcagccag gtgtggtggc ttgcgcctgt agtcccaact
44101  acttgagggg ctgaggcaga ggatcgcttg agccagggag ttcaaggctg cagtgaacca
44161  agattgcacc actgcattcc agtatgggct acagaggaag actttgtctc aaaaaaaaaa
44221  aaaaaaaaaa aaagaaaaga aaaagaaaac atctctcata atcataccac ctagaggtaa
44281  atatgcttaa aactttagtg tttgttcttc caggcttttt ttttttttt ggttgcttgc
44341  tcatacatat acacaaatat atatttttaa tttaatagcc aataatatgc acactcttct
44401  tataatctgt atacctctat tgtcactatg gtttctccac accccacagg atccattcag
44461  gtaccatgag ttgcacttgt gttatatctc tttagtctcc attaatctag aacaattttt
44521  ctgccatttt gtggtgttca tgatgttgac atattgaaga atcttagttt ttataatgtc
44581  atattgaata tctctcagtc tggatatatc tgattgtttt cttgtgatta gattcagttt
44641  aaatttttt ttttttttgt aagagtattg tgctggtgtt tagttcccat tatatcacat
44701  caaaagacgt ataatgtggc tgggtgcggt ggctcatgcc tgtaatccca gcactttggg
44761  aggccgaggt gggtggatta cttgaagtca ggagttcgag accagcttgg tcaacatggt
44821  gaaatcccgt ctctactaaa aatacaaaaa ttagcctgta atcccagctg gttgggaggc
44881  tgaggcacaa gaatagcttg aacctgggag gcggaggtta cagtaagcca agattctgtc
44941  actgcactcc agcctgggca atagagcgag attccgtctc aaaaaaaaaa tgtataatgt
45001  caatttattg cattattggt gatagtttag gtgatgtcaa ccacagacct cttaatttta
45061  aaggtttttt ttttctattt gtaattaatg ataaatctgt ggggtaataa ctttgagact
45121  ctgttaatat gccattctct agcaactttc cacttaatag ttttagcatc cttgtcaaag
45181  tcactgatta caattacact gattacatat gtaggctgca agatggaaat ttgctaattc
45241  caattctaat tcctcctaca tttattagga agcatttact gtgtaaagaa attctcttcc
45301  cctttttttc cttttcagta ttactgtgag ctcatatgtt ataatctatt aactgattgc
45361  tgtttctgtt acttattaca ttgtccaaRt ttgaccagtg agagtccttt aagaaggccc
45421  ttgtgtatac acaaacactt ctctctcttt ctctcactSt ttctctgttt tgtccaccac
45481  ttttctgtct ctacctctaa ataaactgga taaatcccta agatattttt ctacaacttg
45541  ctttgtttca ctcatctgtt aattctgtaa ttcttctcat gtcagtgtat gtgggatatt
45601  tcattctttt taatgataca tagaatggct caaccataat gtacttattt ttttgtttca
45661  agtttttcac ttatatatag tgctgacaaa catgtgccat gtatatatat gtaataattt
45721  ctgtgggata gatacctaga agtaaaaatt ctttttttt ttttttttt tttttgagac
45781  ggagtctcgt tctttcacca ggctacagtg caaaggtgag atctcggctc actgcaacct
45841  ctgcctcctg ggttcaagcg attctcccgc ctcagcctcc tgagtagctg ggactacagg
45901  catgcgccac ctcacccggc tatatttttg tattttgagt agagacgggg tttcaccatg
```

FIGURE 2-M

```
45961  ttggccaggt tggtctcgat ctccagacct catgatctgc ccacctcggc ctcccaaagt
46021  gctgggatta cagggaagta aaatttctta gtcaaaagat aagtgtattt aaacttctaa
46081  aaaccatttc cagattgact accaaaagga ctgtactaat taccaacaat gcacaaacta
46141  cttgataaca ttgcatattt gtaaattttc ttttattctt ttggatataa tctctttatt
46201  acaatgtgca ttcaccttat taagaatgaa gtttggcatc ttgtggccat ttgccacatt
46261  ttgttttctg ttacttgcct gtcattttct tagtcatttc ttttcttttt ttttctggtg
46321  aaaacataca catatattta gaattagcca gctggactca gtttagatga tcccaatttt
46381  gttggcaaga ttcaaagcat tgtaatcagg agccagtcga acatatgcct tcttttctcc
46441  atcaggccga attagggtgt tgacaccttg gccacatcaa tgtcacagag ctccttcaca
46501  gcctgtttga tctggtgctt gttggcttta acttcctcag tgaacacaag catgttgttg
46561  tcttctatct tcttcatggc agactcagtg gtcagcagaa acttgatgat ggtatagtag
46621  tcaagcttgt ttctcctggg ggagctYttc cgaggatatc tgggctgcct ccagagtctc
46681  agtgtcttgg gctgctggaa ggtgggtgac atgcggatct tctttttgg tggctgtgga
46741  caccttcaa cactgccttc ttggccttga aagccttcgg tttggcttca gctttaggag
46801  ggacaggagc ttccttcact ttcggcactg tcttgtgaaa agcatagtca tttcttgtcc
46861  tggcgcggtg gctcacacct ttaatcccag catgttggga ggctgaggca tgtggatcac
46921  ctgaagtcag gagtttgata ccagcctggt taatgtggtg aaactccctt ctctactaaa
46981  aatacaaaaa atttgccagg tgtggtagtg ggcacctgta atcccagcta cttgggaggc
47041  tgaggcagga gaatcgcttg aacccaggag ggcttgaacc caggaggcgg agattgcagt
47101  gagccgagat cgcgccattg cactccagcc cgggtaacaa gagcgaaact ccgtctcaca
47161  aaaaaaaaag aaagaaaaga aaagcttagt tattaggttc tctttttttc tcattgactt
47221  attggaattc ttttttaat ggtctttgtc tgttatatgt gttgcaaata ttttccatcc
47281  ttgcaaatgt tttttatgtt ttccatcctg ttctgtatcc tttaactttg tgatttttta
47341  ttgtttacaa ttttctgtac agttaacacc tgtagatttt ttttcccccc taactctttt
47401  gggactagtt tttaagtttc aaaatgaatt tttatctaaa actcaatcca aagaaaggga
47461  cattttaaa agtgtaataa gtagctagta ttgctagagt cttttgtttt taaatatggt
47521  tgcttacagt gacagacttc catattgtaa ttcagccatg caatttttaa aaaacatggt
47581  taaatatgaa cttcatgaat actttattta gagaaaacta tttgcctatc ataatacatg
47641  aatattaaat aaaatgtgga aaatatagaa cagagagaag aaaaaaacat ttctaattgt
47701  cctcatatcc agacagccat agtattttgt tgcatttctt ttatctcctt ttttttgaga
47761  aatatgctta ttctcaagta attgcataag gatctttaag ttgagatttt gctaaaagat
47821  agtatgccaa atttcgctta ttgaatttat aattgtttac tgtaattact tttcagagtg
47881  ctcacttcat ctaaatctct tttctttttg aaatttgagg gctgcagtgt gtcacttagt
47941  ataaatatct ttgtgggaag atcttctaat attagggtac ctctgaccgt gatgtcactt
48001  atcctgatat cctgattgta gtgtgtctga attcaggaga atgtcaggta tatgttagta
48061  cttgacagaa gttaggtagt gtacttacca gcagttgccc tcaggtttag ttgctattca
48121  ttaaatgaga taaggcttac taagagatgt ttccccctat ttatttttca gtttactgtt
48181  agcctagccc aaggccattt ttgtttttta gtgtatgcaa aaagatatcc ttaggagtgt
48241  tcactggata gtagaatatg ttcaagttac aaccatttgg tactttactt cttttgtgct
48301  ttaggacagg ggtccccaac caggtggtac tggtccttgg cctgttagga actgggcctc
48361  acagcagggg gtgagcggct ggtgagggag cattactgcc tgaactccac ctcctgtcgg
48421  atcagtggca gcattggatt ctcctggaag cgcgaacact attgtgaact gcacatatga
48481  gggatctagg tttcatactt cttatgagaa tgtgatgcct gatgatctga ggtggaacag
48541  tttcatccaa aataatcccc ctgtcccatc catggaaaaa ttgttttcca cgaaaccagc
48601  ccctggtgcc aaaaaggttg gggaccaatg ctttgggata tatttctgac aagtattatg
48661  catgttgtag gtactgtctt tctcgtaaac aatattactc attttaaaat ctttttggc
48721  cgggcgcggt ggctcacgcc tgtaatccca gcactttggg aggctgaggt gggtggatca
48781  cgaggtcagg agatcgagac catcctgact aacatggtga aaccccgtct ctactaaaaa
48841  tacaaaaaaa ttagccgggc atggtggtgg gcgcctgtag tcccagctac tcgggaggct
48901  gaggcaggag aatggcttga acctgggagg tggagcttgc agtgagctga gatcatgcca
48961  ctgtactcca gcctgggcga cagagtgaga ctctgtctca aaaaaaaaaa aaaaaagtct
49021  tttttattga tgagtttcat agtcaaaagt ctaatgatat attaagtaca gaaaaagtgt
49081  tttaagtttt taaatggcaa cttttatata cctgttttat atataccgaa atataattag
49141  gatagatatt attaaacaca tactgcagat gaaataacta agcttcagag atttggtttg
49201  ttcaaagtta cacagctaga agatggaaga ctgaaagtcc actggcccct gtgtgaaaag
49261  ctgaatgggg aacatttcca ctgatttgaa ttatccatac ctttgtacaa ctcaaaacaa
49321  ttgtaaacac atatttttat tttcttaaat atgagttgtg tacattgcat atattagtca
49381  aagattgatg ctaattaaat ggccatagaa acatgtcttc tgtgttaccc attggaaaca
49441  cactgagcac atatgagttt tatgtgttta tacattgttt attttgttta cccatatgaa
49501  aaagaaaata tgtttcccat actcatgttt tatatggaag agagaaagat tttagaacgt
49561  atttagtcac taattaaaat gaattggagc agatatacag tgctagaact ttggtctaga
49621  gctcagtaaa cccaacaggt atgacagaga aatatgggat aagtggcagg taaaagtggc
49681  aagtaaagaa ctctataaag gtaatgacac agtatggatc atttttcagg gacagtgaag
49741  ctgaagaaaa aatacataca gacttgctgg ttgttctttg agcttgcaca acaccattct
```

FIGURE 2-N

```
49801  tccagattac aaaggaagaa tttcctggga atggatttca gaatcattcc taagattaaY
49861  tgtaggaggt ttaaaactct gtgcaaatgt gcacgtttgc tatgtgtgtg gaatattctt
49921  taaaaatata attgaaaatt ttctttattc tgaaaaagtg aactgctgtg atttaccttg
49981  atttggctga atgaagttaa taaattggat agctccttaa gggctttatt ctcttggtaa
50041  ttgcctactt gtgtggatgc acattttatc tagtttgaat ttgtttcaaa ggaacattat
50101  tttgtaaata aaatatatat tttttcccat gtggtattgc atagcacatg gagaaaatag
50161  attgatgata ccgtttcatt ttatgatttt tctttgaata taatgttttg ttcttttatt
50221  ggaaataaaa ctgaccccaa gaataaaatt atatttcctt ttaattttat gatagtgcat
50281  ttaaagtaat gtatatctac ccaaggtgat gaagatgatg attgccaacc acaatttaat
50341  gacattaaag aaattaatga gttttgattc ctttgcttat cttgtgtgtt atagtctaac
50401  tccttttaaa tctgttcttt gacattttat atgactttca gattgtgtgc gtatatatga
50461  tacatataat tatgtattaa agtatgttaa ttaaaaatgt taattaaatg aaattaatta
50521  tatagtttaa tttgtaacta ttaatattaa ctatatttag cataactatt aacatagtat
50581  atatgaaatc catatataca tatatatata cacacacaca tgcataatct gaaaaaatat
50641  atatactttt ttgagacgag attttgctct gttgcccagg ctggttcact ggcaccctcc
50701  acatcccagg ctcaagtgat tctcccacct cagcctttct agcacctgag actaccgtgt
50761  gtgccaccat gcttggctaa actcttttt gtatttttg tagagacagg gccttactat
50821  gttgcccagg ctgatctgaa actcctgggc tcaagcgatc atcctgcctc agcctcccaa
50881  agcattggga ttgcaggcat gaactacacg ccttggctga ctttcagatt ttattaaaat
50941  tcactgttat aaaacctttg aatagcataa ggataacaat taaacagttg cttgaagcag
51001  cttgtgtaag tcatagctgc aaaccaactt tttagttttt tgcttactat taggctattt
51061  tagttatgat ctatatttgg taagtagtgt tttaatgaaa tacttacatg taaccactga
51121  catataatga cgtctagtta attaatgttg gtaatttaaa attatttact aattgtatat
51181  ggctcttttc ctttctcgtg ggtgatttgt ttaactcagt attcattcat ttatatactg
51241  tagttattta ttttagtttg aatttctttg atggttgtga gaggtcacgc tccaatttag
51301  gctcctggtt gtctggttag tgtttaccag ttcaaataat cacatgatct tacattggat
51361  atatataaac ttacccatat ggtgaagtct tttaggagac gatcatgtta aatgtttctc
51421  actgtggctt atttaatgtc atcagtttat tgcgtgttac tttggatgcc ccttagaact
51481  tggcaccaag acctaccctc ttagaataaa cctggtcaaa aggatgttgg gatggttgct
51541  attttatctt tagaaatgac ctattttgag atgtaataaa tgtgtctcag gagactggag
51601  aaatgcccaa acctaaaggt gatacttcag gacaagaatt aagttgcaat gtgttcatta
51661  ttatatatgc tttagattgt ttctttgaag gatgttttac attagggcag tgccttcctt
51721  tttagaggaa gggaaatttg ataataagtt agtgtatgtc cactaccatt tgggaaaaag
51781  aaatctattt aggagtcatg actatcatca gtagagaagt gtacttcatt gttggttatc
51841  agtaagtagt tcagcatggt tttttatagt ctgaatgaaa tactctctca gaacttttca
51901  atttccattt atcagattcc atgtttatta tactttggtg gtagtaataa caaacttta
51961  ggcatcttga tgctgagaaa gttatttatg attatggatg tgacattgct taccagcaga
52021  ttttgttggt ttctaagata ccggcagcct catcaatgct gactgagttc acagacaaaa
52081  ttgagttgta ggaattttaa tgatcaaaac ctgatttacc aagatattgt caatgaagtc
52141  ctgcctgaac ttcttattgc agctttgata aataatctac tcttttaaac tcaaaatcag
52201  ggaaaatact ttaaaaattt taggcagatc tggaagccta aatagaaact ttaaatatac
52261  agtcacccct cagtatacct agggaattgg ttccaggacc cctgcagata ccaaaattag
52321  cacatattca agtcccacag tcggacctgg ggaacttggg ttatgaaaag ttgatgcttt
52381  gtttatgtga gttttgtatc ccacaagtgt attttcactc tgtgttgtgt tgaaagaaat
52441  ctgcatataa gtgcacctga ccagttcaaa cctgtgttgt tccattgtca actgtaactt
52501  gtaattccag agatgtatct aaggcaactt tgcttttgat aattggaggg caattgaatt
52561  gaagtatgac atgtcctact ctaaagtgta aagagcattg ccaagttatc ttgtctgttg
52621  ggtgtgttca tgagaatctt taaggatttt tattacatta gctgcctaaa taatatgatg
52681  aactcctctc agaatgttga gttttgtgag tgataagagg aaaaaagaga atgacatatt
52741  gttttcatcc atttcagtta tagacaaatt ttcacagaaa gaataattta aaatttttga
52801  acagattaaa tacatcatag agttagagat tggctcagtt gtgcatggaa ctcagtctgt
52861  ctaatgttac tactctgaac ttttcaaact gttctatagc tttaggtgca taactttgct
52921  tttacaattg tattatacct gtttgtatta taatctgcca cctccaaaat tagatcaaag
52981  acttttggt ttgtatcttt gttttgcatg agtttgatct tttatatata tattttagta
53041  cctgctacaa agggacctca Rtaagtgttt gccaaccaat aattaatcta aattctttaa
53101  tgacaacttg tatttttat attttatatt tttatttttt atgagatgga attttgctct
53161  tggcacccag gctggagtgc agtggcgcaa tcttggctca ctgcaacctt tgcctccagg
53221  gttcaagtga ttctcctgcc tcggcttctc gagtagctgg gattctgggc gcctgccacc
53281  acgcctggat aatttttgtg tttttaggag agatgaggtt tcaccgtgtt ggccaggctg
53341  gtctcgaact cctcacctca ggtgatctgc ccacctcggc ctcccaaagt gctgggaata
53401  caggcatgag ccactgcacc cagcctaatt tttatatttt ttaaattaaa aatttattgt
53461  tttcccacct ttaatgcttt tacagtagga ctaaaactat ttttaaaatg tcatttattt
53521  cattgtgtat actacattag gtgaaaaagt atataaaatg agtttatttc tgtattttaa
53581  tacgtttctt ataaaccatt ggtgtttcat atctttttcc tcatctcttt gctattcttt
```

FIGURE 2-O

```
53641  ttcttttgtg attttatgtg actttaacaa aataaaccct atttaatcat agtgtttttg
53701  tttttttag gttgttatgg agttgatgga gagagtgact ctattatgag ttcagcttct
53761  gaaaactcca ctgaaccttg gttttgtgat gcctgtaaat gtggtgtttc tcctagctgt
53821  gaactgtgtc ctaatcagga tggaattttc aaggagacag atgctggaag gttaatgtcc
53881  taattatgtt ggttcatatg tttgctttat aatgtgtctg cctttgactt ccctgttctt
53941  ctgtttatga ttacacactt tcatttgatt gcttctttgt atcttcttgg catagaataa
54001  atgaccagtg ggatacagat ggagaagggt gagaggaaag accatttatt gagtacctgc
54061  atttgccttt acatccataa tcttaattaa tcttaataat tgaccctatg atgaggcagg
54121  tggtaacgtt tctcatttta taggtgtata gtacccaggg ctcagattat ttaaataact
54181  tattcgttca agaccacaaa ctggtgaatg gtcaattttg aacctagatc tgtataactt
54241  tatatcagga gatagaactt aatacagtga tttaaaaaga agtttagtat aaagaatttt
54301  taagctgtga taaaagagtc actataaggt ataaaaaaac tctaaaggat gctctgggcc
54361  tgagggagat tatccaagga atgacaaaat tggaatgatg acctcctacc ctaggttggg
54421  gttcaggctt tgttggagaa attatagtta aaccctactg gatggagaga aggccctctg
54481  tgttgagagg gccatggaaa aatgatcttc atgttgggcc agggctgtga ggtcactgag
54541  ggactgtgta tatagctggg gcagagcaca ggaggaagta ctccacccat acctgctgaa
54601  aagccccctcc gttctacgag acctcctgcc attgagttta tagctccctc taccattgag
54661  cttagtgtta tgctcactct gagaagaaat gctgaagaaa ttctgtccat tattgcagag
54721  taggtattga agcatgcgtg agaggcaatt aataagtggc acaaagacct gtgttctttt
54781  tactacattc acactcatgt ttgttaacac atttgagcac ttgtcatcat aacggtgttg
54841  ccatcttcca tgtttttttt ttttttacag gagactacca tctagggttg gtgactaaaa
54901  ttaagagtaa ttttcaacct tttgtatttc gtggagatca cgctctgcct cttcctctcc
54961  aatataagag ggggggtggg tacttaaatt tttattttgt cacataagga tgttttagta
55021  ttttattaaa catgctttat attaaaaaca agattaacag caaagcaacc tatataaatt
55081  atcaccatca tttcccatat tttaacataa aaactagggt atattatgtc ctaaaataaa
55141  ggataggctt ttaaatttaa taaattttac tttagtaaag taaaacaaat tttctcttat
55201  ttttatcatt tttgcctcaa ttggtggact gatccttgag aaatgttgcc atacaagatt
55261  gctgggatag cgtagaatct accattgaat cttggagaat cccatgtgtt gatgttcagg
55321  acaaatgact aaagaatctt acctcttctt ccttattgtc tatccaaaaa taaaatatat
55381  aacaaatgat ttgatgtatt gacatgaatc tgtggtctgt atcataggca taggcattac
55441  tctaagttgg gcctgagagt ggcctagtat tagggagttt agtaagcatt aagtgtgcac
55501  cttcaagttg gactgtctat gtttggattc tagtcttttt tttttaatag ctataatttt
55561  aggcaagtga tgtaaccttt ttatgcctca gtttccttag cagtgttgag aagactaaat
55621  gagatgatcc atgagaaaca actagcagtt tctgtaacat aataagtggt cagtaaatgt
55681  ctccttactc ttactactgc cgttactgct gctactagtg ctttcttatt accattattg
55741  ctccacttat tttaggtaat catggagtat tagctcatgc atagattttt ttttttttt
55801  tttggtaact cttccagtcc tgagaattga aatttttggt gataaattgt ggtgaataag
55861  taaggcagtt ttggaaagac agaacttcct tttagggaag agaatttatt ttactgtttt
55921  tcttgttgat aatttaaaag ttctgttagg acactgtcaa gtagtggtta cttcatgcta
55981  tcataataga caaaacaact ggcaaagttg tttctgaatg aagaaaactg aggcctcttt
56041  accatttagg tttactgagg gatttttttt tggatactgt ggacgttaca taagtagaca
56101  cctcctgtga tggtgcaatc tggcaaaaat tcttttatat actatttgaa cacttgtgct
56161  aatgtctaca cttacctgta ggctgtactg gaaattatag gccttcacaa tatttccaag
56221  gacgtaatac tagtttgatt aaaggtataa aaagccaaga aaattcgaga agccttttag
56281  tatatgagct tcaaatggaa aacagggaaa aagaaagaat attgtatata aagaataaaa
56341  aaggttttga agttcaccat cacatgagat taattagact ttttttcttc tattcttgat
56401  gggagccaca agttctacat agaaaatgat gtgatttaat gtaagaaaac cagtacattc
56461  ttgtcattaa atttttattt tatatcttat ggagagtaag acttataaac ctgactgtgg
56521  gaaatttttt ggcatgctct ccttttttat tttttattta tagttttact atttgtgggt
56581  acatagtaga tatgtatatt catggggtac atgagatgtt ttgatacagg catgcaatgt
56641  gaaataagca catcatggag aatggggtat ccatttctca aacacttgtc ttttgagtta
56701  caaacaatcc agttaccctc tttaagttat ttttaagtgt acaatttagt ttttcttgac
56761  tgtagtcacc ctgttgtgct attaaacagt aggtcttagt ctttccactt tttcgactca
56821  ttaaccatct cccacctccc cccaccattt cacccttcc ctgcactatc cttcccaacg
56881  tctagtaacc atccttctcc tgtctgtgtc catgagttca gttgtttggt ttttagatcc
56941  cacaaataag tgagaacatg tgatgtttgt cttgctgtgc ctggcttatt tcacctaaca
57001  taatgatctc tagttccacc caagttgttg caaatgactg gctctcattc ttttttatgg
57061  cttagtactc cattgtgtat tgtgtaccac aatttcttta tcggattcgt ctgttggtgg
57121  aaacttaggt tgcttccaaa tcttagctat tgtaaacagt gctgcgacaa atataggagt
57181  gcagatatct ctttaatgta ctgatttcct ttctttcgga tgtataccca acagtgcagt
57241  tgctggatta tatggcagct caatttttag gaacctccaa agttttcttc atcatggttg
57301  tactaatWta catttcatgc aacagtgtac aagggttccc ttttctccac attcttgcca
57361  acatttgtta ttgcctgtct tttggatata agctatttta actggggtga gatgatattt
57421  catgatagtt ttgatttgca tttctctgat gatcaaggat tttgagtccc ttttcatatg
```

FIGURE 2-P

```
57481  cctgttttag catatgagct tctttttata tgtctttttt tgagaaatgt ctattcaaat
57541  cttttgtctg tttttaatca gattattaga ttttttcctg tggagttatt tgaactcctt
57601  gtatattctg gttatttatc ccttgttagt tgggtaattt gcaaatattt tctcccattc
57661  tgtgggttgt ctcttcagtt tgatgatgta ttctttgccc agcttgttgt gatcctgctt
57721  tttaaaaaaa cctgatatga tatctttcgt ctgtttttgc ttcggttgct tgtgcttgtg
57781  gagtttcgct caagaaattt tacccagagc aatgtcctgg attccccaat gttttcttgt
57841  agtattttca tagtttgagg tcttagattt aagtctttaa tccattttga tttgattttt
57901  gtatatggca agagatggcg tctagtttcc ttcttctgca tatggatatg cagtttttcc
57961  agcaccattt gttgaagaaa ctgtcttttc ccatatatat gttcctggca cctttgttag
58021  gcactaagtt cacttaggtg cgtggatttg tttctgagtt ctgtattttg ttcctttgtt
58081  ctgtgtattt gttttatgac agtaccatgc tgttttggct actgtaggtt tgtggtataa
58141  tttgaagtca ggtagtatga ttctttttttg cttaggatag ctttggctct tctgggtctt
58201  ttgtggttcc gtatacattt taggattatt ttttctattt ctgtgaagaa tgtcattggt
58261  attttgatag ggattgcatt gaatctgtag attgctttgg gtagtatggg cattttaaca
58321  atattgattc ttccagtttg tgaacatgga atatttttcc atttttttggt gtcgtcttca
58381  atttctttca tcagtgtttt atagttgtta ttatagaaat cttttacttc tttaagttaa
58441  ttcctaggta tttaatctta tgtgtggcta ttgtaaatgc cattactttt aaaatttctt
58501  tttcagatta ttcactgttg tcacgtagac atgctcctga tttttgtatg ttgcttttgt
58561  cttttgcaac tttactgaat ttatcggttc taatagtttt cttatgaggt ctttaggttt
58621  ttccaaatac aacattatat tacctgctaa caagaataat ttgacttctt tctttccaat
58681  ttggatgccc tttttatctt tctgttgtct gattgccgga agctagggct tccagtacta
58741  tgttgaataa cagtggtgac agtgggcctg cttgtcgtgt tccagatctt agaggaaaag
58801  cttacagttt ttcccattca gtacgatact agctgtgggt ctgtcatata tcgcttttat
58861  tatgttgagg tatgtctttc tatcccgttt ttggggagtt tttatgatga agggatgttg
58921  aattttatta aatgcttttt cagcatcagt tgaaatgatc atatggtttt tatccttcat
58981  cctgttgatg tgatatatca tgttgattga cttgcttatg ttgaaccatg cttgcaaccc
59041  agggataaat cctacttggt cataatgatc tttctaacgt attgttaaat tcggtttgct
59101  ggtattttgt tgaggatctt tgcgtcagtg ttcatcagag atactggcct gtagttttct
59161  ttttttgatg tgcctttggc tggttttggt atcagagtaa cactggcctc atagaatgag
59221  tttggaaata ttccctcttc ctctgttttt tggaatagtt tgagtaagat tggtattagt
59281  tcttctttaa gtgtttggta aaattcagca gcgaagccat caggtcaagc agtcccaggc
59341  ttttctttac cgggagactt tattatggct tcaatctcgt tacttgttat tggtctgtac
59401  aggctttgga tttcttcatg gttcaatctt ggtaggttgt atgtgtctag gagttcatcc
59461  atttctttta gattttctaa tttattggca tatagttgct catagtagct gctaatgatc
59521  ctttgaattt ctgcaatatc agttataatg tctccttttt tatttctaat tttatttact
59581  tgtatcttct ctctttttc ctcagcctgg ctaaaggttt gtcagtgttg tttacctttt
59641  caaaatacta acattttgtt tcattgatct tttgtatatg ttttttattt caattttatt
59701  tatttctgct ctgatcttta ttaatctact aattttgggt ttggtttgct cttgctttcc
59761  tagttcttta agatatgtca ttggattgtt tatttgaagt tttttctctt ttgatgtagg
59821  tacctatagc tataaacctc cctcttagta gggcttttgc tgtatcccat agattttggt
59881  atgttgtgtt tccattacta ttcgtcatac tcttcttctt tatgtccttt gttgggtgtc
59941  ctcccagctg ttctgaaatg aaatatggct ttgaactaaa ttccttaatt tttaaaaata
60001  agctttatgt tgcaatagaa ggcagcattg aaatctttat gtctcagttc tttctttttt
60061  tttttccaat cagacggggc ctggctctgt tccccaggct aaagtgcagt ggcacgatat
60121  tggctcactg taacctctgc ctcctgggct caggtgatcc tcccactgaa gcttcctgag
60181  tagtgcaggt tataggtgtg cgccaccaca ccaggctaat gtttgtattt ttctgtagag
60241  acgaggtttc aacctgttgc ccaggctggt cttgagctcc tgggctcaag cagtcctccc
60301  acctcagcct tccaaactgt taggattact gtacctggcc aatttttttct ttttttctgg
60361  agagaacgtc tggctttgtc accctgactg gagtgcagtg gcatgatctt ggctcattgt
60421  aaaccccccgc ctcctaggct caagcagtcc tcctacctca gcctcccaag tagctgggac
60481  tacaggggca tgccacgaca cctggctaaa ttttgtattt tttgtagcga tggggtttca
60541  ccatgttgcg caggctggtc tcaaactcct gagcttaagt gatcctccta cctctgcctc
60601  ccaaagtgct gggattacag gcatgagcca ctgtgcctga ccacttcttt cttttttttt
60661  ctttaaaatt ctagatttag taagtcttag ttctgtccct tgctagactt gttatatttg
60721  aaaatcagtt aatttttctg aaaccaattc ttcatatgta aatgaacaac ctcctactca
60781  acagaagctt ttaggctttt ttttccttgt agttaacatt tcttgagtga agacagtgag
60841  ccacactctg ctgtaagcac ttcacataca tctttataac ttccaaacta cctttatggt
60901  gctctgtgcc tagtagacac ttataattga ttactgaatg aaggaatgaa aataacgtac
60961  ttgacgtgaa tttgttttac atgctgtaaa ggcattactg atacagagca cttcaaactc
61021  caattttatt gattaatata gttttcattc ttaaatgagt cttttgagaa gtatcactct
61081  acttagtctg ctttattaaa atatatgaat ttacaataat ctgttaatag tttgtgctag
61141  gatgacaatt ctgactgttt tttggtggtt ttatgtgatg cttatttcct gttcctttcc
61201  ccagtcagca tctgtcataa tagctccaag caaaattaat ggtaaagtaa caaggaagat
61261  tttttttaa attgtagatt tgtcttaatt acaaagttat aaataatgtg catctaaaca
```

FIGURE 2-Q

```
61321  gatgagcaag aaactattta cagctggtat taaaagggac catagttgtc tcggccatga
61381  ttttcattga ctttgatagg tatttttaatg aggactcaag ttggtatcct cacatgagtg
61441  gttgggagaa tgatatactt gtatgccagt cacattgaaa tggcagttta aatctgtagt
61501  acttgtgtac tacattttaa gaaggtcagt atgaagcaag cagccagctg ttaggtttat
61561  caRagataaa ctcctacaga gtatagcgac atttaataca tcagtggagc atgaaaaaac
61621  tgaacaaatt aaattgagaa agcagctgct tgttgaagaY gtgtatcttt atattgaact
61681  ccataggatg tatttgggag gttggattaa aacattctat caagaaaatc tcaaaacctt
61741  ggctttttc cccaataaaa aagataattt tttaaaaaga tttagaacta tttttcccag
61801  tgcctctaga gaataatttt tggtcttagt cctctaaaac tcataatgca ccacaaaaac
61861  cagaaccatg tagcatgcat agcctctgaa gttgtgataa agcagaattt gttgttaata
61921  tattcacagg gttaatagag atttgctgct gtgaccctct ttcttacaaa agtgctaaga
61981  tctactgttt aaatttcaaa gtcagtgatg ccacattttg ttttgttgta agcaagcaga
62041  aagagtgcct aatatattaa aaaaggttca gagcagtttt taaaaaatga aattagaaac
62101  aactgcaatt tttttcagtg tactgaaact ctgagcattc tttgttttag tcttatagca
62161  tcattcaaac tttaataaat atttatattt ggggcatata aattttagta tRtagattgt
62221  atggtaaaca gataagatga ataaaaatgt gaattgctta ttctgtacct tttcccagca
62281  gttttaatct ggaaaggaag ttaaaactgt atacaaaaat tgtaaaatat gtccttaaag
62341  attattaccc cagcagcttt tggaaaagag taataagaga aattttacct aattactgca
62401  gagtaaattc acagttatta aattagtatc cttaattctc aaagggtttg aaataatgaa
62461  cgtgactacc tttccttgca tatgaaaatg atgttttgtt tcatttgaat agtagtttta
62521  tgcaaatatt taagaataga taagtgcaaa gtgcagtgag tgaattttaa gtgaaaggca
62581  attatagttg ttagctctca gttttggcta agctgctgga aaagtaattt cagtcccctt
62641  ttttgtaact aagtgttatt ctgtaggtgt ttcatatttt ttaatgatgt gcctgagatt
62701  gcgagatttt aacatttgat tacagaattc tcttggtagt attttgagat ggcaagcgtt
62761  ttatatgttt gttttgtgtg tgggaatttt attaaaagat ttgatagcaa aatcatatct
62821  tctcttctta gatgggttca tattgtttgt gccctgtatg ttcctggagt agcctttgga
62881  gatattgaca aattacgacc agtaacacta acggaaatga actattccaa atatggtgcc
62941  aaggtgagac acaaagcatt tttgattgct tgaaagagaa aggttttact tgttagttta
63001  cctggctttt tatttgtatt atgttgactg cagaaatgct tactaggaat cattctattt
63061  agcactaaat catttattat tgatagtagt tttcctagaa agccaataat gcataaattc
63121  ttctatacat tcttaattta gagtaacgta tgaggtaaaa tgataattgg tactactacc
63181  atttccctgg cagtgttcac aatgcctaat acatagacca aaaccttatt tacaattaga
63241  tggatcaaat ataagcctaa tttctgactt tacttttgaa gtggaatttg tcaaaattat
63301  gaaaaagaat attagaagtg atctttgatg ctgttactgt aattgttttg gggcgtgccg
63361  ttaacatgcc cttaatagat ggtgaactta attgatatat gttatgtgtg ttctgactgc
63421  ttcaccaacc tggtgttccc ctatctttct ccctgacttg gttggggcct ctctgttcgc
63481  tgagacacag tgaaattgaa atttggccaa ttagggctgg gtgctcacgc ctgttatccc
63541  agcacttcgg gaggccaagg tgggcagatc acctgagatt aggagttcga gaccagcctg
63601  gccaacatgg cgaaacctcg tctctactaa aaatacaaca aattagctgg gcgtggtggc
63661  gcgtgcctgt aatcccagcc actcaggagg ctgaggcagg agaatcgctt gaacccagga
63721  ggtggaggtt gcagtgagct gagattgcgc cacggcactc cagcctgggc gacagagtga
63781  gactccatct caaaaaaaaa gaaaaagaaa tttggccaat taataattct acagtggcct
63841  ctaagtgttc aagtgaaagg aagagtcata catctgtcac tttaaattaa aagctagaaa
63901  tgattaagct tagtgaggaa ggcatgtcga aagcctaggt aggccaaaag ctgggcctct
63961  tctgccgaat agccagtatg tcagtgcaaa ggaaaagctc ttgaaggaaa ttgaaagtgc
64021  tactccagca aacacatgaa tgataagaaa gtgaagctgc cttattgctg ataatggaga
64081  aaattttaat agtagggata gatcatagca gccacaacat tcccttaagc caaaacctca
64141  tccagagcat gaccctaact gttcaatatt gtgaaggctg agagaggtga ggaagctgca
64201  gaagaaaggt tgaaaggtag cagaggttgg ttcatgaggt ttaaggaaag aagccatctc
64261  cgtaatataa aagatgaaga agatgcagca ggtactgatg gagaagcatc agcaagttat
64321  tcagaagatt tagctaagac aaatgatgaa gatttgtaca ttcaacaaag attttcaatg
64381  tagatgaaac agtcttctat tggaagaaga tgccatctag gacttaccat agctagagag
64441  gagaagtcac tatctggctt ctaagggcag gctgtctctc ttgtttagag gtgaatgcgg
64501  ctggtgactt taagttgaag ccagtgctcc tatactattg tgaaaatcct agagccctta
64561  agaattatgt taaatctact ctgcctgtgc tctgtaaatg gaacaacaaa gcctgtgcac
64621  agcacatctg tgtaccgcat ggtttactgg atattttaag ccccttgttg agacttactg
64681  ctcagaaaaa agattccttt caaaagatta ctgcttatcg acagtgcacc tggtggtcac
64741  ccaagagctc tgatgaagta caaggagatt aatgtgtttt cctgcctgct gtcacaacag
64801  ccattctgaa gcccatggat caaggagtca ttttgacttt aaagtctgat tatttaagaa
64861  atacatttca taaggctata gctgccatac atagtgattc ctctgatgga tctgggcaaa
64921  gtaaattgaa aaccttctag aaaggggtca ctttttaga tgtcattaag aatatttatg
64981  attcatggga ggaggtcaaa atatccatat taacaggagt ttggaagaag ttgattgcag
65041  ccttcatgga taactttgaa ggggttgaag aattcagtgg aggaagtcac tgcagatgtg
65101  gcagaaataa caagataact agaattagaa gtgtagccaa aggtgactga actgctgcga
```

FIGURE 2-R

65161 tctcatgata aaaatttgaa cagatgagaa gttacttctc atggatgagt gaagaaagtg
65221 gtttctcgag ttggaatcta ctcctggtga agatgctatg aacattgtta aaattacaac
65281 aaaagactta ggatattaca taaacttgat tgattaagca ggaagagggt ttgagaggat
65341 tgactccaat tttgaaggaa gttctaccat gggtaaaatg ctgtcaaata gcatcacgtg
65401 ccacagagaa atctttcata aaaggaagga tgaatctatg cagcaaactt cattgtttta
65461 ttttaagaac ttgccagccg ggcacggtgg ctcacacctg taatcccagc actttgggag
65521 gccgaggtgg gtggatcacg agatcaggag atcaagacca tccccgggtaa cacggtgaaa
65581 ccccgtctct cctaaaaaaa tacaaaaaat tagccgggca tggtagcagg tgcctgtagt
65641 ctcagctact tgggaggctg aggcaggaga atggcgtgaa cccgggaggc agagcttgca
65701 gtgagccgag atcgtgccat tgccctccag cctaggtgac agagcgagac tccgtctcaa
65761 aaaaaaagaa aaagaaaaag aaagaaattg ccaggctggg tgcggtggct cacgcctgta
65821 atcccgacac tttgggaggc cgagttgggt gggtcacctg aggtcaggag ttcgagacca
65881 gcctggccaa agtggcaaaa ccctgtatct actaaaaata caaaaattag ctgggcgtgg
65941 tggcacgtgc ctgtaacccc agctactcag gaagctgagg caggagaatc acttgaaccc
66001 agtaggtgta ggttacagtg agccgagatt gcaccactgc actccagcct gggtgacagt
66061 gtgagactcc atcaaaaaaa aaaaaaaaaa agagagagaa attgccacag ccactccaac
66121 tttcagcaac caccaccacc cccatccgtc agtggccatc cacatcaaag caacatcttt
66181 caccagcaaa aagattatga cttgctgaag gctcaagtga tcattaacat tttttagcaa
66241 taaagtattc ttaaattaat gtacttaaca ttttaaaaat acataatgtt attatgcatt
66301 taatagacta cagatagtat aaacctaact tgaaaatаca tggagaaact aaaaaattct
66361 cgtgatttgt tttattgtga tactttattg aagtggtctg gagctgaact tgggttatct
66421 ctgatgtatg cttgtaaatt taacagaggt tatttgagca ttaacagatg tatttattgg
66481 gcagcactta gaactagaaa aggctcagag atgcccactt caatagtgtg agcagtgggc
66541 ttctattggc tgaatgtgga accaaagtaa ggaaattact taattgacta cagctgtgtt
66601 gtttttctgt atgggtatga tttagtagaa agtgtccagt tacgttatct gatgtctagt
66661 tggctgttgt ggttagttga agcttgtttt ttttttttt ttaaatttaa tttattacaa
66721 ggaatttttc taaattaaat ttcaccttgt gtaaagttcc atgtacagaa acaggtccag
66781 gctaatttcc tcttgcttat tttgctttaa gaaaacatca ctgggaaact gagttgaaaa
66841 ttaggtactc tctggcttct ctgcaaaatc agatacaatg ttgtaaatca tgactgggac
66901 agccttctag acctcattgt ctactcattt tccaatctga attatgttct cctgtgttct
66961 cagttgctta gttcactagc attttaaatg attcaggcga ttgaggaaga ttctgttttc
67021 tttgatagat tgttttatag tctaaaaagg ctgtcaggaa tattttcctt atattcactc
67081 tgaaatcccK tgctcaattt caacatatta catgaaatta tccagataat ctaaacaata
67141 tgtctctgtt cacagtgtgg acatccttta aataatttaa tttctcttta tctctctYgc
67201 tcaatctcat gtttagattg ctatttgcca gtatattata ccatttatgt tctactgtta
67261 ttagagaaat tatattaact aggcagtagg atgtcaggat ttgactattg aagtttttaa
67321 gttttccttg tatttgtttc cttcttatgt agttcacttt gaatacttat attaaacaat
67381 tctttaattg tccattgtca gtctttcata atttaacatc tttttctttt atgaagattt
67441 ctgttaccag agtagtttaa acttttagac tagctaattc ccccttttтc ttttaagaga
67501 tagttaaagg aataaaaaaa taaataaatc tgagaggatt tgcggtaatt tcaacagttc
67561 ataccttggt tttataaatt ttctcttttt atacagaaag aatgtacata attgactgtg
67621 ttgtatatta ctaccaattt cacctgatta ttaacctaaa attactttta gttgtcttaa
67681 agatatttaa aacatgcatg atttctaggt tgtaatgcct agatttttтg ttggattttc
67741 tctggtctct tgtttccttt tccaagcaat tttatattgt ttaagaatta gtatagcttg
67801 atggtgggct ttaaagtaag attgtcctgg tttgaagcct tacatttact ttctttgtta
67861 tttttggcaa gcagatttct taacttgtct ttactttaat ttcctcatgt gtaaaatggg
67921 aatagttata gtatctcttc ttcatagtag ttgtgtttaa tgattacata tatttataca
67981 tatagtcctt gacttatgat gggtttacat cctgataaat ccattgtaag tttaaaatat
68041 tttaagttga gattaattta atacacctaa cttactgagc acagcttagt gttgcctacc
68101 tttaatatgc ttggaacatt tacattagcc tacacttgaa cacaatcatc tgacacaaaa
68161 ctttatttta taataaaatg ttaactatct catgtaattt attgaatacc ctactgaaag
68221 tggaaaacag aatggtcgta tgggtactga aagtgtggat tctgctgagc gtgtattgtt
68281 tttgcaacat tgtaaagtta aaaaattaga agtggaactg ttgtaaactg gggactgtct
68341 gtatatgtaa agtgcttaga atagtgactg gcacatagta agtcctcaaa taattttaat
68401 tatggttgta ttattattta ttatcattaa attagttgaa ttttgattta tttgaaccat
68461 ataggtaatt acgtttttтc atataaactc ttggctcgcc tatgattatt gtttgtattg
68521 gtacataatt ggttcattat ttgtatttag cattttaaat atttagtgga cactttgtat
68581 gagaatgcag tctttagtct ggaaataagt ttgctttgag aatttttctg ttactttgtt
68641 gtaggagtgt agcttttgtg aagaccctcg ctttgctaga actggggttt gcattagctg
68701 tgatgcaggg atgtgcagag cctatttcca tgtgacctgt gctcaaaagg aaggtctgct
68761 ttcagaggca gcggcggaag aggtaggttt atttaaaccc atagttggtg aacatgttca
68821 caagatatct tttgactcta tgtcctataa taagtgtaaa taataaaaga aattttaaca
68881 tttattctta aagttcttgc caagatcact gttctaaaac tttaagagta aatgcctatg
68941 agttttgtca gttttaaagc tagtcagacg ttcagaaaac ataatttтta tagtgagatc

FIGURE 2-S

```
69001  tgacagataa aatgaaagat aatccaagtt gtcttaaagc attaaaatgt cggcatttca
69061  gccccatgaa ttatagttaa aaaaaaatag catattttat acgagaagct gaaactgtat
69121  ttgggtaacc ttgctctatt tttctttaaa taaaatcatt actgtgtctt tagtatctct
69181  gttactatct ctttactgtt acttttaatt ccataatatt cctttcagat ttattggctt
69241  tacgttaaaa aatcagttgt tttcagtgtt tggggactgg ttaagattcg ttcaagtatt
69301  ctgagtttaa aactattttc attacactat aaaaatgtta cttacttttc tgaactctgt
69361  tctttcatga gtattagtgg agacttccag agattacata cagtataaca gaacccaccg
69421  aatatagaaa caattttgag aatccagcta ccttttttta tgtcagttat ctagacatta
69481  gtatgtaaaa tgtaaaacag tgccactctc ctcactcaat ttttgtgtgt atgtgaattg
69541  tatttagtaa aagtgtatac ttcggtaatg gacccattat ttaaaataaa ttaatgaata
69601  ttgcaaaaaa ttttcagttt taatttttaa tagggtaact gtctaaaata taactcacat
69661  aaaccctttg gagtcctcag taatttttaa gaatgttaac aaattgtgag caggttcagt
69721  tcattcctta agcttacata gtaatgaatg tttatttagt aatttttatt gcacattcat
69781  gatggcttta tggtaatagt ctaagcctct ccttgttatg ttttttcttt tattgggtaa
69841  attttgttta ttgtgctgag attttatctt agtacattta tttcatttcc agaaagtgag
69901  caaagatact ctgatgaaga aaaattgtag aaaaaaattt gaaggatata ttttatgaat
69961  aaacagatct tataggaatt gtatttttga tacatgtagc tgtttttgat acatatagct
70021  gttccaatct cttcatttat atagctgttc cagtctcttg attttaagga Rttgtacttt
70081  gatgaaggct actagatgaa cagtgagcat cttctcagct actaggtaga catagagtta
70141  taattatacg gtaaaggtag tctggattcc tcatagctac aactcaggat Rgtgggcaaa
70201  agtacgagag ggcaatacag gtaagagtta gaggaccaga taaatactac gtaaatactt
70261  ggatcaggga atctatagga taaggttagt tgtctacttc tggaaaactt tgagctttgc
70321  ggttagtagg gtgcttttgg agccagtgga gagctttctt ttttactatg gtagtttgga
70381  ttcctttgca cagcatcata tgtcatcatt tttggcaaga tggcagctaa ttaacaaaaa
70441  tgtgtgaagc agccatgttt gggacagtgg ggtgtaatgg agttatggga tttggagagt
70501  gcatctccgt ctaaaggcat gcaaattaaa tgaaactaaa cataatatta agctacatag
70561  tttacattct gtggtctgtg tatcttacct ataagctaat ttgaggaaat ttacagggaa
70621  acagatacat taggatgtga gaattagccc tgaggcatct tccttgcctt gctgtgtccc
70681  ttgattaatt aggtgggagc cctggatgct ggaataataa aatgcaactt ttctccctat
70741  tcactacctt actcagtgtg ctgttgtttt ggggggaaat gatttaaagc atatttacat
70801  atatatgtta ctggtttctt gttataaaca gagaattaga aaattctaaa tgttttctga
70861  acgatatttc aaaacccaga atattgcatg aatcaatatt aatgaagttt gcatttgaca
70921  gtaatcattc tgtcttgtct atttctagat ggttgtactt atgtacataa ggatttatgt
70981  aatgtaatag catttaaaat taagaaacac attttaaact aaaatagatc tatgtttgtg
71041  agcatatctt gccttctatt tccattttat actctttaga cattctaggg aacttttcaa
71101  aagactatat caagttggat ttactacttt ttcatggtgg ttttatatag gctgttaaaa
71161  aaaaaaggtt gaaaaataga ataaacatat cagtttcatc taatgacatt gtgtaccttt
71221  ctgtattttt tatgactttg tatagccttc tgacatagtc ctttatctca ttggtgagaa
71281  tccatgtgac atttctattc tttttcatct actttgtgaa tattccagat ttattaagaa
71341  tagtcaaagc atgttatgac acaacatgct ttattagaag gatatttaaa accatgttat
71401  gacacaacat cttatttta aaactggact atagggctgg gtgcagtggc tcatgcctgt
71461  aatctcaaca ctttgggagg ccaaggccag aggattgctt gaggccagga gtttgaggcc
71521  agtttgggca acatagcagg accccatctc taccaaaaaa aaaaaaaaaa aattagctgg
71581  gtgtgatgtt gcctgtagtc tgagctacta gagaggccga ggtgagagga tcacatatac
71641  ccgggagttt gaggctgcat tgaactatgt tctcaccact gcactcagcc tggagaatag
71701  agtgagaccc tacctctaaa aattttaaat aaataagtaa aaaaataccg agctattgtg
71761  caggctcagt tatgctgact acaacttatg ttactatgtt ttacattata tagtctaaag
71821  ttacttatag tccaagtaag ccatatggta gtttcttctg taattaaagt ataatttggt
71881  gctattgaac attgctgtac tctgttctgt agaaaattat atttgaaaca atataaaacc
71941  gggtgtggtg gcacacgcct gtaatcccag tactttggga ggccaagcca aggaggatct
72001  cttgaggctg gcctcaagag gattttgaga tgagcctggg caacatagca agacccaatc
72061  tctacacaaa atgtaaaaat tagccaggtg tggtggtgtg tgcctatagt cctaactgct
72121  caggaggctg aggtgggagg attgctgagc ccagaagttc aagatttcat tctattagag
72181  catatatggc tgaggtgggg tggacaagaa tattgataat gcataattaa gaagataagt
72241  ttgatatgcc attaggttat tattgtattt aaatgtttct ggtttcagta aacacatcaa
72301  ttttactatt aagataccag taaagcgtag gtacgttctt caaatacact gtttttacaa
72361  gtgccaggta aggttctcaa tgtacaagac cgaagttgtg tcttctttca gtttatataa
72421  ctttgatctg ctatgtcagt tttacccagg gagctttgtt taaaaaaaaa aaaaagattt
72481  taagattcta gaccaaatca tggcatctag aggaccacaa gctaaactta atccatagat
72541  ggattttgtt tgttaggagt agagttaaaa aagaaaacct tgaatttaaa tgactttaga
72601  tggggcattc cagtttgcca caggctctaa gttctcacag gccccagcat ttactaactg
72661  cttctcatca ttcatttttg tcttatattt ttctggcatc ttttaggtct ccttcaccca
72721  tttatgttga ctattactgc ttctcaaggc atttgtattt aaggcccttg cctcagctct
72781  gctaagtcag aatttgcaga gttgggctct agttacctct atttagaaat ctccttaaat
```

FIGURE 2-T

```
72841   gaatctgatg acagtcaggt aatggttaag gaagatggat cttacatact ggacataggt
72901   cttggttcag caacttaacc tctctggctt tcagttttct atccataaaa ttgatattaa
72961   aatggtagaa cccacctcat gtggtagtgg tgaaggttaa agagatagta tatacaaaag
73021   cgcatgtgct tatttcatgg tagatattca gaaacatgtt tttccctttc tgtcttccac
73081   tgggagtagg atcttgaaag ctccatttct ccaagcaggt tgaagggata ttccatgtca
73141   accaatgcat taattaggac attattttga ctggtttaac aaaaccaagt aacgagttgc
73201   ttaaacgaag aagttgaatt cactctaata tacaattcca ggcggattgg atactgaggg
73261   gagtttggtt ccagggatca gttcccttct ttcttcttgc ttcaccttcc tttagtctat
73321   tgcactctcc tacacagtca aagatggttt atcaccatct tgactggatt cttcctcatt
73381   ggaagtagca aagaggaggt taagagtagg cagcttcctt tcaagggcat ggtctgggca
73441   tttcatacaa cttattggcc agaacttaat cacatggcca catctagctg tggggaggct
73501   ggatcagata gtttctagca gggtagccat atgcccagta aaaactcata ggggtctagt
73561   actaatggga agaatggata ttgtaagaca attagtagtt tctgccatag taagtattgg
73621   tgtttttctt aagtcagagt cttggtccaa atgcttaact cacctatgca agcatgaatc
73681   ctatctctaa tgactcaatt tcaacctcat ttactttgaa acattagaat aattatctag
73741   atctgggtat ttgggctgag gttcggtata atctttattt ccatttaaca tctctgtcac
73801   caagatttag agcatgtcat gagaaagcag ctgtttgtat gtgttaatgt agtctcctaa
73861   ttaatttttg tatttaattt ttaaagctat atgaagtaat tgtaagtaaa ccctgtttta
73921   ctaatgtggt aactgaggcc cagagatgtt aagcaatttt ttgaagtccc atagcctgta
73981   aatgatagag ccaagatttg tatccRggtt tttatgtctt caaattctgt gatacacatt
74041   gctgtactgt gacatctata gatatcacac tacagaattg cctcatatgc ttttttttta
74101   atattaagga cacattagat tttttttttct ttctgttctt atatgtctct cctcggtagt
74161   gttgtcattt ctctttgggt tgtaaaactc tacttagaga cacttttgac ataacttaga
74221   gaaaaactaa tagcttatgt ttactgttcc tttttgggac agacataaag aatgaacagc
74281   ctgttgtttt gttagcaatt acattgtaaa ttacttttttt gtactacatc tttacaagtt
74341   gtttttgata gaataaatat tctttttttgt acactacttt ttatgaatga aaatgtactg
74401   ttaggctatg agagagctgt aattccactt gagtttttag ggaaagagtt aagggcataa
74461   cttaaatttt tttattgaca gggcttagaa tttcttaatt ctatttttttt ttttttttttt
74521   ttttttttttt gagacagact gtcgctctgt cacccaggct ggagtgcagt ggcgcgatct
74581   cggttaatta caacatctcc ctcctgtatt caagcaattc tcctacctca gcctcctgag
74641   tagctgggat tacaggcaca tgccactgtg cctggctaat ttttgtactt ttagtagaga
74701   caggcttttg ccgtgttggc caggctagtc tcgaactccc aacctcaagt gatctgcttg
74761   cctcggtctc ccaaagtgct gggattacag gcatgtgcca ccacgcctgg cccttaattc
74821   tatttttatc ctcagtagaa tctgtccatt gttagccagt tagaaatgcc acacttatct
74881   tgttggctag ttgaatgttg ttgttttgca atattagttt atgcataact ttaatattta
74941   atgattattt ctacttaaga ttattaaaat cctatttgtt ttaaacagtg acttttttaaa
75001   aaaatacagc cggccctccg tgtctgtgga ttacacattc ttagattcaa ccaactgcag
75061   atcgaaaatg tttgaaaaaa aaaaagcgta ataaaaaaga atacaaatat aaaaccaata
75121   cagtgtaaca actatttgca catttacatt tgtgtattag gtattataag taatgtagag
75181   actgcttaaa atatgtggga gaatgcatgt aaactagatg caaataccac accatttaac
75241   ataagacgct tgagcatcca cagattttgg tatccttggg gggtcctgga accagtcccc
75301   catggatatt gaagggctac tatacaagta cacccaaata atacattata tttgatttca
75361   atatctttaa acctgcaggc cctttaaaga aaaacgttgt tggtaatttg aagttgacat
75421   acaagatcaa gtttctaaag aagtaatttt tgtcattatg tacaaaaatt atttttgtaa
75481   tttattagat gtaactagca ctagtagatt ctttgaaaag gaagtaatat tcttttgtgt
75541   taggtctttt atgactcttg ggagtacaaa tgttcactat ttttctgtgc tttctttgtR
75601   taggatatag cagatccatt ctttgcttat tgtaagcaac atgcagatag gttagacaga
75661   aagtggaaga gaaaaaacta cttggctcta cagtcctatt gtaaaatgtc tttgcaagag
75721   agagagaagc aactatcacc agaagcacag gtatgggatt catgtcaaaa cccgtatgtt
75781   tttgttttaa ggttatgtaa ggattttacg tctcgtgtgg cttccagtga catgtgacgt
75841   ataataggaa gtttggttag ttacctagga aatggattgt gattaaggta aatagtgagt
75901   ataatttttg gctgaataca gtactcctca tttagtcaaa cagagcagtc ctttataaag
75961   taaattttca cattgaaaaa tgtgactgtg aggaaattac tctatcagtg caatatattt
76021   attacatctt gtttcagttt gttcaaagga tttagtcagt ttgcctatat ttaaaatata
76081   attttatata cattaaacct atggagacaa tagcagttat taagaagaga aacttcagct
76141   taattagttt ctcaaaactc tgattataag ttaggacagt tgttactagg ggcagatata
76201   gtttattgta actcaaaaat tatacagtat accttaagga tcattcatac ctataatttg
76261   ctttggtgtg agaagttctg attatctttc tgtgaacatg ggaataaaaa tcaatgtaaa
76321   aatcttaagg aacatgtttc attttattat cttttaaaat ttgaatacat ttcttttata
76381   ggcaaggatc aatgcccggc ttcagcagta tcgtgccaaa gcagaactag ctcgatctac
76441   cagaccccag gcctgggttc caagggaaaa attgcccaga ccactcacca gcagtgcttc
76501   agctattcgt aaacttatgc ggaaagcaga actcatgggg atcagtacag atatctttcc
76561   agtggacaat tcagatacta gttctagtgt ggatggaagg agaaaacata agcaaccagc
76621   tctcactgca gattttgtga attattattt tggtcagtat agacactggt acatgcacat
```

FIGURE 2-U

76681 tttcactgag aacaatttgt taatgaaaat gtttgatata ctgttaattt ttagaaattt
76741 attagccatg ctgcatatat catagaggga tgtttatttt cctaatatgt tgtgggtttt
76801 tacaaaatgc taggttcatt tatttgtaaa tatacttaaa agttttaaac tatgtaggtt
76861 tatcaatgtg atcatcttta tagctagttt cttcctatgt cagtttttac ttcctactaa
76921 agtaaaattc gatatttcat ttaaatagag agaaatatgc gcatgattca aattcaggaa
76981 aatatggctg aacaaaagaa tataaaagat aaattagaga atgaacaaga aaagcttcat
77041 gtagaatata ataaggtaag ttagctacaa aatatgcaac ataatgtgat agatcttact
77101 gatgatttct tttgtgaatg tatggcattt cttttatgtt ttgtttttaa tttcttctgg
77161 agctctttgg ctcttcccta taactcctac tcctcattag cttcattatc atcactgtct
77221 cttactgagc actttgtgtg ccaggcataa ttttgatctc tttacatgta ttatttaatt
77281 cccagcttaa ctttatagtg taggttataa tcattgcttc cattttcagg gagaaaactg
77341 aatcctgaag aagttaaata atttgcccag gttcacacaa gtagtcaagt cacagacttg
77401 acaaatgaag tcagttggtg tgaggccaac tcatgtgact ctgttctcta tttctaatca
77461 ataaacattg ctatagagct tcactttcta agtttgtaga aaatttagaa acttttcttt
77521 aaattattat ttatcttgaa ttatattata atgtttggaa taaaaatttc caaatttgaa
77581 ttttcatgta cctcttcatc gtaatgagaa aagattcaaa aagaagtatc ttataaaagc
77641 aaaatgtggt taaaatgtta taccaagtat aacttgacat ataaaacggg tattgactta
77701 ttgttagaaa acattttaag tcataaaaaa atagtaggaa gtacgtttcc aataatcagc
77761 tacactgtgt agtaaaaaga gtacttggaa ttagattcag aggacctggg gattcttgct
77821 agttctgtga tcttggccaa ctcacaaaaa ctctatgatc cttagttttt ttcattggga
77881 aaatagtgga aatacatctt gagagagtca caaggattag ttagcagaac atgatgttat
77941 ataacaaact gtaaaagaag gtataaacgt agttaccatt attagtaaaa atgccagcag
78001 aaaacaaaat gcttacagca acaaatactg ctgaataatt ggatagcctt gtctctcagt
78061 ggtattgtgg cttgaatata ttaaaaatat atggaggccg agcgcagtgg ctcatgcctg
78121 taatcccagc actttgggag gccaaggcag gtggatcacg aggtcaagag atcgagacca
78181 tcctaggagt ttgagaccag cctggccaac atggcaaaac cccgtctcta ctaaaaatac
78241 aaaagttagc tgggcgtggt ggagggtgcc tgtaatccca gctactcgag aggctgaggc
78301 aggagaattg cttgaaccca ggaggtggtg gttgcagtga cacaagatcg tgctactgca
78361 ctctagcctg ggcgacagag tgagactcca tctcaaaaaa aaaaaaaaaa aaatagatcg
78421 agaccatcct ggccaacatg gtgaaaccct gtctctacta aaaatacaaa aattagctgg
78481 atgtggtggt gcacacctgt agtcccacct actcgggaaa ctgcagcagg agaatcgctt
78541 gaaccccaga ggcggaggtt gtagtgagcc gaaatcgtgt caccgcactc cagcctggtg
78601 acagagtgag actctgtctc aaaaaaaaaa aaaattatat atatatatgt ggaatagaaa
78661 tgtagatatt tcttaaattg tcagatattt taatgaagtt tactaatttg gcttactttg
78721 tagctatgtg aatctttaga agaactacaa aacctgaatg gaaaacttcg aagtgaagga
78781 caaggaatat gggctttact aggcagaatc acagggcagg ttagtttctt tccaattgct
78841 gtctccttca gttcttgctc cctatatgat tttcaaagaa cagatttttg atacaactgt
78901 catttgacta accgaagaat tctgtttgat cagggcctaa tcctttggga taatgatttc
78961 tgcttgaaat acacctagag cattggaagc tgcctactga gaacctctag tttaatttgg
79021 ttgtttcaac agctatcttg tgtgcaaaat tttttcattt gctctaactg attgacctgt
79081 catttgtgat gaagcttaag acttgaagca tgagtttttc ctactacatt cttctttcct
79141 ttaatttagc agttttagat tccctgccac attttttttct ttcctgtcct tatgctgtat
79201 ttttttttct attctttgga tgtataattt aaaattaaaa aaggcaaact agagatattc
79261 atttttatgt taatttggtt ttcagctgta taagagatgg ttttttaaac tttaatgtta
79321 aaatgtattt ttccccttta acagtagctg tgatcgtagt tattaaaaaa aactttattc
79381 atgaacacct ttagtctgtt aaagtttttg gtgttgaata attattgtta gatggatttc
79441 tttgaatgcc ttttttagta ctgtgtatat aattgataaa ttacagtttt cgcaaatatt
79501 ttttattttt taaggatgat ggagacccac tgtttgtggt cctttcacta aaaagtagtt
79561 ttatttgtga aggagctttt cgaacagata tttttcatta tacttgtttt agaggtttat
79621 ggagagtcaa attcatcaca cagcttattt aaatatttta agaagtgata catactgatt
79681 tgtactgtgc tataccagct ttgaagatgc tttatgtgaa aagaatgtgt gtggcttgta
79741 tcctaaagaa tgttttaaaa ggtgagaatt agtagtcgcc tctgggagga tcagcctttg
79801 ttcctgttaa taggttggtt gatggatttt tggaagatta tgttttatac ttaatattta
79861 gaggagttaa tgttgagtat aaggctttag tttatttgcc aagctagcat tatcttttgg
79921 cttgatttgc aagttttttct cataagaatc tcctgtttaa aatactatct accataaaag
79981 acattttcat ttattaaaag acttttaaca gaggtttgca aatttccttt tgcaaatgta
80041 aaatacacag cagaagtaga attcatttat actgcttgca tgacactttt tttttcagtt
80101 ggttgaacag atgtgggccc accaggccac tgagaaatga ctttgataga tggttgctgt
80161 ctgagaagcc agcatactac agattagtag aaacccaaag agaatctcac ttgctttgtt
80221 ggagagtttg gtccaataag ataatttttc taattaacct atggaacacc aatatagctg
80281 gggtgctgaa gaaaataaat ggcctcaata ttctttttgt caagtgtact gtgttttgtt
80341 ttttttaaac ctcctacgta tgtgtctttg tctttgaaag ctgaaataga agtttaattt
80401 ttatgtttta gtgtaccact gaaaattgat gaaagtctgt aattatagat ttctaatgga
80461 aataactatt agaaaaaagc tcattttaat ccttttgcc aatgtaaaat ctgataaagc

FIGURE 2-V

```
80521  agatttcccc ctaacaatcc tactatattt acttggcttg agtggcaatt agaggaaaat
80581  gttgctttta tttctttctt aaaacaatat atactacatt tgttcaaatc aatagaagtt
80641  gaatataccg gcaattttgc gagcacccaa ggagagaaaa ccaagtaaaa aagaaggagg
80701  cacacaaaag acatctactc ttcctgcagt actttatagg caagtaatga aattaataat
80761  gatagaataa tgttgtgtat ttttttaaac tgatacttat tagaaagaag atcctgcaaa
80821  tatttgagaa tgaattatgt ccataccaga atttaagtgc atttgaagtt cattttctca
80881  aaagatgcat tccagtttga atgatagagg agatttttgt tgctgtttta acctaaactc
80941  tatggaaatg agcagtaccc tttggggaac tggaaaatta ctttagataa cattatagtt
81001  attgtcttaa ataatttaaa caatataaaa gatgagcatt tatattttat tttacttgaa
81061  tgaattgtag agtttgtttt ttagatagtg atgggtatca tatatgtata tatatagaat
81121  gaatctattg cctttgtagg cagtggataa tgaagataag ttttagtttt gagtttctac
81181  actttttttc tcttgatgct cagttgcaat cactttgttc gcataagtat ttttgttttt
81241  tactttcttt gctcaatatt ttcggtgttc accattatag taattttatt atagtcaaga
81301  cttctttcca taaataaata ggttacgatc tcacttttttc ccaaataatt atattaagtg
81361  ggagaagttt agaagatgct ggtgtttttg tgtgtgtgtg tgtgtgtatg tatgtgtgtg
81421  tatatttatc tatgtatatg catacatata tgttatgcat gtgtaaatac ataaatattc
81481  ttctcaaaat ttctttctca gaaatagtag caataacaac aatatcaatc tccataacta
81541  ccttctgttt ttttgttgga ggctccctct acaagtaata gcaaataact tccttgaagt
81601  gtgggaaaag cctacattgt cttagtgtac aatgtgaaca ttagcacatt ctcaggccag
81661  ctttttcaaa cacggttggt tcatttggaa ccatcaggca atatcactat gcttctgact
81721  tttatgacta caattcattt ttgtttcaga atatatttta agtgattttt ctttttttact
81781  gagaccaggt actggttatt tatttattta ttttggcttt ttaagtgtgt gtgtgtgtgt
81841  gtgacttttt tctaggagta tttaaaaaaa aaacctaacc atcattgctt tctgtacttt
81901  ccaaatataa atatacttta gattttcagg tgatggtgct tgtgtacgca ggtgtagtta
81961  aattgccttt ttgcataaga gaagttagat ttctgagaaa ccaaaagatc ctaaaataat
82021  agtagcaatc acaaaattcc acaatctttg caataaatat caattgatgt gtgcgttttt
82081  aatttttta cttttgaatt taattagtat atttttaaaa cgcaaaattc agtattgaaa
82141  agagactcat gtatacttgc agtatatcta catgttgaag actgtttgaa aggtttacaa
82201  ctaatggtag gaaataaact accttctgtt tgtgaagtaa tttttttata gttttagaaa
82261  attggaaaat tactttgttc tataatttat gtgtctagag aaaaaatgaa ctaatgtgaa
82321  atgggtttta agtaataact aaaattaaaa tgaatatgat gttgaggata aatacttat
82381  atatgtaatt ctttagaagg tttgactcta gtagtattta aatttttacc agatccaagc
82441  tatttttaac attctcaaaa gtatttagaa ctaacagtta ctgacattgt aaaaagatgt
82501  gttaatgaaa gaaattgtct atctcaacag aagaaatagt aagttaaatc aagatgaaat
82561  tttgtaagtt gaaaaatatg ctataagtaa actgtttcac tatgataatt attattgaaa
82621  tctttacttg ctgcctttat taaaagttgt gggatttgta agaagaacca tgatcagcat
82681  cttcttttat tgtgtgatac ctgtaaacta cattaccatc ttggatgtct ggatcctcct
82741  cttacaagga tgccaagaaa gaccaaaaac agttattggt gagtaaaata gaggagattt
82801  agatgtttga attgatttac tcttagtttc agcagtataa gatgaaatac tatatttgtt
82861  cataataaag tatttggcac attttaaaat tgactgtcat caatatttat tatagcatta
82921  tgtgactaaa tctataacag ctttttgag aaatgagttt ttcattaatt gtggctatta
82981  tcttagcaat agattcttct ttaaggcatg aattctgtcc attgcttata attctcttgt
83041  tttattgttc agaatttctc aaatgcttta ttgttatttt gaaattgaat ttatctaaaa
83101  tttgccatca tattatgctt ttacttgcaa tttaatattt tatttttgga ttagagtcaa
83161  aggctttact tctggtattt ggtatttgtg ctagtcattt aaattcaacc attctttgac
83221  atctagaaac atttgataaa catttctgat tttttttttt attaacttag gcatatcgaa
83281  tcactacatt tgataaaggt tatagtatct ataatgagtt ttatcctttt cttaatttta
83341  ctccaagaaa gggacagatg ttcagatttt aatagaatga tttctcatag gaaggagaat
83401  ataggaaaag gaaaactttg gaatgaacat gaatattctt tttatacaag ttacctttcc
83461  agttttagaa aagagtgaac cccaaatttt taaatggtaa tctctacaag aggttgtgtg
83521  tgttcctcaa atacattagg catatttttt gacctcattg ttttatcact taaggctacc
83581  tctgttcatt tgattaggaa gctgaaattt tttcactatt ttcacatgat tgctttttta
83641  tagtgttctt tgtatcagtt tgttttgaga tattcatact ggcctgttct tttgaaatgc
83701  agtggaatgt atttattatg gtttttatgt gcttcagaaa ctgctatagc atattcaaat
83761  cccaacaatg ttaataagca tcctgggtga gtttttaaat tgaaagctag tttgatagtc
83821  tccagagcca ggatgttttg ccatgaactt atatagagtt atgcagttct ctccttaaat
83881  gaagttatgg accctcctga tactgctcag cctgcagagc tctggtttcc agcatgtgag
83941  agtttctgca gctaaaaaaa aacacacaca aaaaaaggga aatcatgttc tttgtagcaa
84001  caaggaggca gctggaggcc attatactca gtgaattaac ccaggaaaag aaaaccaaat
84061  actgcatgtt ctcacttaca ggtggaagct aaacactgga tagtcatgga cataaagatg
84121  gggcctacta gaatggggag ggaggggagc aaggtttgaa aaactgttgg gtagtatact
84181  cactacctgg gtgataggga tcattcatat tccaaacctc agtatcatgc agtataccca
84241  tgtaacaaac ctgtgcatat acccccctgat ctaaaataaa agttgaaatt aaaaaaaaaa
84301  gaaaagaagt aatgtacata gcctgaggtg ttttagtgag agcacttcta aaagggcgta
```

FIGURE 2-W

```
84361  aacaacttag agccttttca ttgttgttag tggtttttaa gtaaaagtag aagttaatgt
84421  tgtagcacag tgaagccttt aatttctcct tgtttctagt gataacattt cctttggcct
84481  ctaagatcaa cttagctttg gttagcagtt aactgggtgt attaattttg gagtgtgcaa
84541  gattcagaga attgtttttt ctttagtgta cctcaaatgt gggaggcagt ttttactttt
84601  tgctttttaa cattttcagt ttatttttga agttatacta aagataataa atactcaaaa
84661  cttaacactt tctaaatatt ttgtcacatt ttactattat ctatgctcta gaggttattt
84721  aagtctattg catctggatg gtagaaaaaa aatgataaac gcaggttgag tatcccttat
84781  ccaaagtgct ggggaccaga agtgatttgg attttgtttt tgtatttgtt ttggaatatt
84841  tgcattgtac ctactggtcg agtatcccaa atctgaaaat ctgaaatctg caatgctcca
84901  atgaacattt tctttgactg taatgtcagt gctcaaaaaa catttggatt ttggagcact
84961  tcagattttg gatttttgga tttgggatgc tcaacctatg atagttgcta ctacacaccc
85021  ttttccccat cctacattct gtgatgccat gttgataatt tgaaattgaa attgccatag
85081  ttaggagtat ttgtactgta gaaattacta aatactacat aacagggctt tcccccagag
85141  ggtcaagtgt tagacattta ctagcatatc actgactgat gtctgcttta gactctacct
85201  tgttcttact agtcctttca ttagggaaag ggtcctttaa tttctccttg tttctagtga
85261  tcaaatttcc tttgatcact agggtcctgt tactggactc ttatttcatg actgtgcctt
85321  agtctgcagc ttgaacttcc acaggtccct ctggtataat atttctgaat ggccggttta
85381  tttttgctgt ctttatgttt tgcaagcctg tggtagcacc cttactacta aataccctgt
85441  aagcaaactg tacctagaat taattactat acaatctgtt gccttctcag atatcatatt
85501  ctactctaat aactatttta agcacttgct atccagcctt ttcttccagt accagcactg
85561  ttttcaagga gacactagtt ctagatttgg atgctttctt cagtccatcc cagagttctg
85621  tggactgtgc tgggcaggtt aagacatgtg acatcaaaga aaagatggat acagatacag
85681  aagtaattat tggtggaagg tagggaagtt gagtgaattc accttggaca ggctcagata
85741  tttcagtgat ataggagtct Rttgagaggg aagagacctg ggatgatgat gggctcttga
85801  gaagacagca gatggttggt aacaacttac tgtggaaaaa tcaaactgac ttgctgaggc
85861  tgtgtagcaa gaatgtttag tgcataagta aatgtgatta tttgattttt cttgaattat
85921  gctctgctgc ccaggagaag gagaggggaa aacctagggg cttcttaggt tgggccaggg
85981  atgtgaaaag ctagtaaaga atattattaa taggtgggtg aagaattcca cagagaatag
86041  agaggctagt gatggtgtga tctaggttct attttgtttg tacctctttc tagtatttaa
86101  aatgcatttt gtctcatttt tagtttatgt tattttgatt tttgttgttg tttttaaaaa
86161  tgtacaatct ccacactgcc ttaaattcta agacacattt ggtaggatac aaatggtaag
86221  ggcttgattt attagagtag ttcaaatagt taatcatctc ttaattttag aagactaatg
86281  ttatatttct gcctccactc atctgattgg aacttgtgtt gccttttcca attgtttgaa
86341  gaatattctg aaacacatta tttaacagag aattgagtaa tatatccttt tattccttgg
86401  atatgtcttt attttatgat gaaaagaaaa ctcaaaccct tctaggaata gtttacagag
86461  ggtttttttcc tttgtttttc ttgataattt gttcatttac atttttttat tgcctcttct
86521  tcccattaac cgctagtagg aattccagca ttagaatggg acatttactg tatatgaaat
86581  aacaggtata tatctgctat taaagttttt acaggtatta taactacaaa aagaataaaa
86641  tttacaagtt gagttaaatg ttgttttctg cccttcaaag agatacccat aatgcagaag
86701  taggagggac tgaggtatta cttgtctgag aagaatgata tttgagccac ctcttaaagt
86761  tttatgtagg cacctgtcaa gactttcctc taaatacagg agaaaacatg gctaacggct
86821  tagacataag aaaaaaaaaa actctggtgt gggatcctga gttcttgatg actgaaaaag
86881  ataaattgtg gaggagaaaa aagtaaatca tgaattattt ggaatatgga atgttactgg
86941  atttatttat tgcattattg catttaaaga ttcataaaat ccttttttttt tttaacttac
87001  acgaataaaa ttctctaaat tagtccactt tctttttttt ttctttttttt gagatggagt
87061  ctcgctctgt cacccaggct ggggtgcagt ggcacaatcg caactcaccg caacttctgc
87121  ctcctgggtg caagcaattc tcctgcctca gcctcccgag tacctgggac tacaggcgcg
87181  tgccaccacg cccagctaat tttttgtatt tttagtagag atggggtttc accgtgttag
87241  ccagaatggt ctcgatctcc tgacctcatg atccgcctgc cttggcctcc caaagtgctg
87301  gattacaggc gtgagctact gtgcctggct taaattagta cactttcttg tctaagcact
87361  attaaacatt tttctttagt aactaagatt ctgaaattcc attgggtcat cattctgatt
87421  ataggttgta gttgtatggt gctagaaaca tatagaaggt atagaaaatt gagaaaaggc
87481  cagacgcagt ggctcatgtc tgtaatccca gcactttggg aggctgaggc gggcggatca
87541  cttgaggtca ggagttcaag accaacatga tgaaaccctg tctctactaa aaatacaaaa
87601  attagctagg aatggtgatg gggattcgta atccctgcta ctcaggaggc tgaggtagag
87661  aattgcttaa acctgggagg ctgaggctac aatgagctga gattgcgcca gtgctctcca
87721  atctgggtga cagagtgaga ctctgtctca caaaaagaca gaagaagagg agggagggca
87781  ggagggaggg aaagaaggaa ggaaagaagg gagggaggga gggcgggaaa aagagaggaa
87841  ggaagggagg aaagaaggga gggaggaagg aaggaaagga agggaaggga gagggagggg
87901  gagggagagg gagggagggg gagggaggaa gggaaaagag aaaaacctct tggtatttca
87961  acatgaaagt tattctgttg cataaattcc tgacatggat ttttattata ctttaaattt
88021  ttaattggct tggatcaatt aaaaggttaa gattatcttc agtattgtcc atgggaatta
88081  taaattttag gactctattc tcatacagtg tgctaagtaa aaatatatag aattaaggtg
88141  agtttagcta aattaataga taatagcagt ttagtggatt attgaaagaa gccaggtatg
```

FIGURE 2-X

```
88201  tatttgactt taggaaagat aatcatgatt ttccttaaaa tattatttga tttcagctgg
88261  gcgtggtggc tcacgcctga aatctcagca ctttggggag ggcggggcag gtggatcacg
88321  aggtcaggag ttcaagacca gcctggccaa catggtgaaa ccccatctgt actaaaaata
88381  caaaaattag ccaggcatgg tggcaggcac ctgtaatccc agctactcgg gaggctgagg
88441  cagagaattg cttgaacccg ggagatggag gttacagtga gccttagatc atgccactgc
88501  actccagcct tggcaatagt gcaagactct gtctgaaaac aaagaaaaaa taattatttg
88561  atttcattga attagacagg ccatggcttt attctaggtt attatttctt accatcttgt
88621  gtaccaatta atgtagcaga aacatttttt acatatctta aaggatattt ataagatgaa
88681  gactatagtg tttaatagat tgttgaagaa tggctaagaa gtattatgtc cataatactc
88741  ttgtagtcta ctcagtgtgg actctatcta ctgcaacaat catgtggcgg ttgtcttgaa
88801  ataggaatgg atatattctg cattggcagt tacatttgtt ttgatttttt gagttgttcc
88861  gtgttttgtg gtgttcaggg tgtaacattt ctattttctg ggaattagcc aaatttcttt
88921  gtcacctaac tattatttat agtggggatg gaagtgaaca taagaataaa gaatgtcctg
88981  acaagggtaa ttgctaccaa aagagaagcg aagaggaaat atttcagttt ttaaaacaaa
89041  acaaaacaat gatttatctg aaaaggtacc tgcttctata cctaggggat ttagataaca
89101  ttatgtttgc ctgttattta agtagacaaa caacatattt taaaagtaaa ttagagcctg
89161  gcgcggtggc tcacacctgt aatcccagca ctttggaagg ctgaggcggg aggatcacct
89221  gaggtcagga gtttgagacc agcttggcca acatatagtg aaaccctgtc tctactaaaa
89281  aatccaaaat taactgtgcg tggtggcaca cacctgtagt cccagctact taggaagctg
89341  aggcaggaga atcgcttgaa tggggaaggt ggaggttgca gtgagccgag atcgtgccac
89401  tggactccag cctgggcgac ttagcaagac tgtctctcaa aaaaaaaaaa aaaaaaagtg
89461  taaattagaa ctgtttgaga aactgactac atttcactca actgatagaa atgatcagtt
89521  agcattaaga ttacaattta atgctgaaaa gtgaagcagt gtgccttttt gtaatataat
89581  atgataattt ctctagtcat ctgttttgaa ggctgttgca tttattatta gttaactaac
89641  aatatctaac atttctgggg ttttcacagt tctgctaata aagtcagaat atacttatca
89701  tttcaattct gagattctgt aaggtgtgta ttttcttttg ccagtccttg atgttttcat
89761  atgggacaaa tgcaggagca atttaatata ttcgtttagt tgtgtctttg ttaatgatgg
89821  taatctgaca acctagggga taacctgcca gctatattta catgaagtgt agcaatagat
89881  acataaaagt tttctgagta atacatgatt atgtacattt ttgttttgct ttcactgatc
89941  agtatttcta gtgctcaaag ttttgctaac ctattcttaa ataattggaa attttaagat
90001  atttaaatta atacgttaag gattaatttg tattttagta ataatatgtt tgcattaaca
90061  aaacaaatgg tacttggtat ttaactcttg ttggagttca aaatttaaag ctgtcacttt
90121  gacaattaaa tattagttta ataagttcac agtattagtg aaagatgtga attttgatgg
90181  aacgaaatca aatcataaat ctttgaaaat acttgctatg agaattataa gctacattca
90241  aagatactga tataaacata taaactaaat atcaaataca attttacaga ttataatcat
90301  ttttattaag agaatactat caaacttttg gagaatttgc ttttgatcat tcttttacat
90361  ttgaagaaaa tatgagtaaa aatttacttt taacaactgg ttttaacttt ttgaaatggt
90421  cctttatgat tgacaatata aagatttcca gcagttgttg tttagatttc tgctgtctaa
90481  atgttagcca ctagccacat atggctattt aaatttaaat caattaatat taaaatgaga
90541  aattttctca gtggtatttg ccacatttca agtgctcagt aaccatctat ggctagtggc
90601  cactgttctg gacagcacag atatgtagat aatcctaata ttataaaaag ttctgttgga
90661  taccacttgt ctaaagtttt agtttcaatt aatttttaat attacccata ttaaaattct
90721  ttataaatgc tactttaaca aattcctttg ataaataaat gtatatttct tcttttagaa
90781  gtagtagggg ccaccaaMta atatgacctt ttttttcccc caacacttct tggagtctgt
90841  ttgtatagtt tctttaagta gtatattcag cacaagttgt caYtgactac aattaaaaga
90901  tatgataact gtgacaagga gtagtgttag ggtccttgaa ggactttgat tttcaggtat
90961  cacagtatat gtcatttaaa gaaaagtcat gctttccaaa ggtaaattat tattattatt
91021  taagacaggg tcttgctcat gttacccagg ctggagtgca gtgttgagat catagctcac
91081  tgcagcctcc acctcctggg ctcaagtgac tctttcactt cagcttcctg agtagctggg
91141  attataggcg tgtgccacca agcctagtta attttctttt ttttttgtg aagacagtct
91201  ctcactatgt tgtgagcagt catcctgcct cagtctcctg aagtgctggg attatgggWg
91261  tgagccactg cagctggcaa aaccaccttt ttaatcagtg ttgaatctca tgttggttca
91321  cttaaacata tagcattgtt aaatgtgttc ttatacttac aggatacatt aaattgcctt
91381  aaatattata ttcatagtag tcactcttag aggagtattt ctgaatgttt tttcattatg
91441  taccctataa tcacattacc tgtgtatatt tttatatatt attcttatat acctggattt
91501  tgtatttata aaaactatag ctaaagccag aagataataa taaatgcatg acttaaattt
91561  ttccccttta tgtaggcagt gctcggaatg tgaccaggca gggagcagtg acatggaagc
91621  agatatggcc atggaaaccc taccagatgg aaccaaacga tcaaggaggc agattaagga
91681  accagtgaaa tttgttccac aggatgtgcc accagaaccc aagaagattc cgataagaaa
91741  cacggtagtt tatttttat ttatcataag catcatacaa ttctgaggcc aaaatttaag
91801  agagtgagaa agacacaggg caaacatata ctcagaagtc aaagaaaaag acatctattg
91861  gttattctta acaatttttt ttctatatta tgaaatatac cacatgctta cctgagtgtt
91921  tgtagtttat attttgttgg aggcataata taaggaacac accagagcat ccatcaccca
91981  ccttgaaaga cagaccattc ccaaagcttt agaagtttcc cgtaaacccc tcgacaagct
```

FIGURE 2-Y

```
92041  catctttatc tgtctagccc agaggccacc tctttactca gctttacttt tctcattcac
92101  ttctttttct ttatagaaga gatttacact ctaaatttct gtacttataa ataatacatg
92161  gtttattttt gtaggttttg aatttcatat gaatggaatc aaatgttttt ctctggtaat
92221  ttgatatttt gaccaaaatt atgttgattt gattcagaca tgctgatcca tgtagttgta
92281  gtttatttat tgtcactaat gtatagtttt ctatggaaag aatagattat tacctagttt
92341  ttcttttgct ttgttgatgg gcagttaggt ttttttcatg tttttctatt ctatacagct
92401  ttgtttaaaa tattcctata aatatctcct gcagcatggt aacaaggagt ttgtgtatgt
92461  attaaggtgt agatttgcta agttacagac tatctgcatc ttcagttttg ctagataact
92521  gcaattgttt tccaaaccag ttcacactca taccagcaat ggataatgtt tctgttgctt
92581  cacaaagttt tgtccagtgt tttaatgctt tccagtgtgg tacagggaaa atagtatctt
92641  gacattttat gttgcatttc tctcagtact aatgattttg ttcatctatt tacatatttt
92701  tggtaaattg tatttccttt tctgggaaat gtgtatgcaa gcctttcacc agtttttctg
92761  ttggggtgtt tacctttttc tcattgcaaa ttctttgagg ccaggtttta aagagtttct
92821  tctggggaaa tcatttatgt tcatgtttac caggtacctt gaaggggcct ctaccaactt
92881  gagaatattt taaataaaag tttggtgtgt atgttttcta ttacatcagg agtgtaaatt
92941  tcaaaccttt aagtgctggc ttacagttat gaattcttag ggcatacttt ccttgcttga
93001  gctgaggcag gcacttttca aaaatcttcc tttccacaag ggagatacaa agatattttt
93061  agggttagtt gtcatcattt gttctgttac ctgtttcatt ttaaagagtc tttaattttt
93121  tccttaaagt aacagttttc tcaaaaaaca ttttcttaat ctgtattaac ctcctatatt
93181  ataatcactt aaaagatctt attcaggaga ttcagcttga ttttttaaat cagttttgtg
93241  ttatggggtc atataaaatt tatgaggcaa atctttatct cttagataat tgaaacagtt
93301  tattttgggt gctctaagta gatcaacttg aaatagtttt tatttgtagt ttcaactgtt
93361  tataactctt tttttaggct tcaaaaatat attttgttta taagggaact agagagactt
93421  ttcctttgtt aaataagccg gctagtttct atttctaaaa tattttctat aattgtttga
93481  aaatggtaac tagttattct taagatctta gagaagagtt attacttcgt atgtctgtct
93541  acaaagatta tcctcttatg aaacattatt tttccagttt ctttattgtt tcaagaaagc
93601  cttttgacta gtattatcat ccttaacttt ggagttatca tatagaggga aattatgtat
93661  attacagtga aatttttgtg cccacagttc agtattactt gagataaaag agaactcttt
93721  aacattaaag gatccttcta tttttttctc acctttactg gtctgctttt tttatcctag
93781  aattataagg aagctttaaa atttgctaat ataatcccat tcctttcatc ttgttctttg
93841  tctctagttt tcttgtttaa aagttttaga cttcttacag attttggtca gtgaaacacc
93901  agggaaagag tccatgacca gtaatttcag gaaatgcagt gtgattattt tcttacttgt
93961  agaaattcca atgcatgttg aagctccaaa aaaaaaaaat cctaatattt agaaacagga
94021  ttacttttat ctattctttt tcttaacttg tttttatcaa aaaagcattt tttaagacac
94081  ttttttttttg gtagcataga gtactgtgtg gtgctaagtt tggaaatgat gctccggaga
94141  ccatggatta cttgcattag gaccatcttg ggtatttgtt gaaatgtaga ctttggaccc
94201  accctagagc tagtatagac tttggaccca ttctggaatt aatcaattag aatttagggg
94261  tttgactaag cagatttctt aggtgctacg cttatactaa ataaaagttg aaaataacta
94321  ctaatggtga ttacctttcc tttttccctc tcatctttc tgcctgcctt ccaacaatcc
94381  attcagaaaa ccagacaatt ggtgcgaggt atcatactaa actgttcaag gaatgtagga
94441  atgagagata cagaagggag tgatcagtca ttttggtctt gatcttacca tttatgctga
94501  gttcctgttg ttaccttaag tttttactta tttctatttg agttgtattt atctttttta
94561  atcaataggt ggtatggtga aaaatgaggt gttctaattg tgtttcttag aacagtgtta
94621  attatgagag accaggcata caacagctta ctaacttgta taaatacctt ttgcaatata
94681  taactaattt ttctctccca ttgttacatt tttaaactat tgctcccaat atcctcctac
94741  ttacactttt atttcagcac aggtattaca attgggaaca ttctgtcttg taatacatta
94801  ttatagtatt tttatttttg tcagtttttc atttttcctg ccataaattg atcaaagtat
94861  aatttgtttt catgttagtc ttattattct gtcatgattt tgaagaaaat tttaccttca
94921  tgacatacgt agggcaatca tgacataccc caagctcatg gacctgaact cctggcagac
94981  agcaactgcc ctgtttacca cagaagaatg ggtcccattc ttcttatgtt tcatcagtga
95041  aactatagaa ttattgtaaa ggcagcctga gatagttctg tttaagtact ttgttattta
95101  gacattttct ctgcattatt ttctatgtgg aaacttttac ttacttttag gcaaggaaat
95161  taattttgaa gaaaacataa attagcttac ccatatccct tttgaagtta ctttaaaatt
95221  ggtgtttata cttgtggaca tattttctaa tactctaatg taacttgaaa gctttcagat
95281  tttctttttg ctttctttgc actctaagtc atgtagtgac ttttcatctt ttgtagaagt
95341  gccaaaggac ttgttgggac ctctctttcc attgttttta cacacacgca tgcatcatct
95401  actctctttg aagttStgag atctttcttc ctgacttaca aattcKggtt gtgttttata
95461  tattttcttt tttgcaaaat aaatgcaata aataggcaaa atacttttta gcagctttga
95521  ctgaaaacca tattgttttt caggatgcat tatttctttt agatcattat tttagaagga
95581  gatttaaaca catctaatcg tttttccatc gcacttaaca atgtgagcca ctagccagtt
95641  ctaagtggca gtctttattt gagagataat caaatatcct agagcctact tttattacat
95701  ttatgtttta actaatttta cgcatataca aagtacctat ttaataatgt ttaagcttaa
95761  gtcttgagaa tctgaaaata tgaaatatgt ttcatataaa tatttttag atgttagttt
95821  aggaaaaatg ttgaagcatt tggattcagt tctactgaag tgactaaagg ttactgtaca
```

FIGURE 2-Z

```
95881  ttatgttcat atttatttta cttattctta tactaactgt ttgaattact acagttctgg
95941  gaaagaagag ttatttgtgg gggctttaca taccagagtt ttctattatc agactcaagg
96001  tgacctctga ggtacgagtc agattataaa acctctatct gtcatagatt cttagaagaa
96061  accctggcaa acagtttttg tagtgaacta gaactcactt tggctacttg gaagagatct
96121  actctgtggg tgtccttagc tcaagtaatc ttattcagaa ccctgagact cctgttttgc
96181  tttttgcctc ttggaaatcc atcactttta tttattccca ggagtatgaa ataaagataa
96241  gaataggtgg agctttcaag actttcctta ttttgtatat accattatct ctgagaaggt
96301  ttttatagca gcacttactt gtcatgtaga atacatattt tattatatat cattagaccc
96361  atagttaKtt cagtttatag tagttaagac aaattggtta tgattttctt tttattctcc
96421  catatatttt cataaccctg ttaacataag ctaaattaga taaaaagaaa ctctacagtc
96481  aattgaccda agggaaagca ctcacttttg gtgactgcca ttccattggt tgtttattgg
96541  tagccaacag aaacagatga caccttgttc ataatttgtt ttttgtatat agcaattttc
96601  tttgaatatt tcatgaactt taacttgttt tcaatgcagt ttcataattg aaagacaaat
96661  atttttagga attatgtata tgtataattt tatatttttt agaaattata tttttattat
96721  atattgctac atataatata tgctatacat ataattttat attcttagga attaaatata
96781  tattatattt ttatatatta gaataaattt tatattgaag cattttgaa tagctgccag
96841  aaagctactg gcatttattc cccagcataa atctaatgct atttagctta acagaggttt
96901  tcaaagtttg acttaattgt cctaattaac attgattttg gaattttgcc catgaataag
96961  catgttctat ttttacatat aagttgcaga gggaagcatt tcttatgatt caccatatgt
97021  gacttacttt aattattaat ttgtataaaY attgatatgt caacaaaaac acaagtgtta
97081  aatttagtga cctggtcaca agtgaatatg tgaagcctag tttactgata tcaaagatgt
97141  taaggtactg actcttttag ttttaaattt agttcatttg ccaaatgaat catgcatttg
97201  acttgattgc aaattaaaat aacctcagct ctaaagaatt aattaaaata cattacatgt
97261  tttttagtcc aaatgataga aaagttagag aaatgtttaa ttatttgttt tagatgaata
97321  aactatttat ttacttattt ttatttttat ttttttgaga cRgagtcttg ctctgtcgcc
97381  caggctggag tgcagtggtg tgaccttggc tcactgcaac ctccgcctcc caggatcgag
97441  cgattctcat gcctcagcct cctgggtagc tgggattaca ggtgtgcacc accacgtccg
97501  gctgagtttt gtattttagt agagatgaga tttcgccatg ttggccaggc tggtttcaaa
97561  ctcgttacct caggtgatct acccgcctcg gcctcccaaa gtactaagat cacaggcctg
97621  agccactgtK cccggcctga ataaactatt taaaagttgc ctgctagata agataatttt
97681  acacctttтc agtttaaata cattgtctct aataccatgc caatctcttc tatggatttt
97741  ttaatcacct cttttcaagt aagttgatca cggacagatt acgagcaagg tgatttaagc
97801  agctcaggtt gtaattgttc cctagctaaa tcaagttctt aaaaaaaaga aaaacaaaaa
97861  attggaatgt gtcaagattt ggaatgagtt ttaaactttc atttacтttt aataggttag
97921  ctaattactg tcaaaattaa tcagtttgga attgcaccct tgcttgatta atcatgtgga
97981  atttccagRt aacgtatctg tgttacattc taaagcacat tcttgaaaag taaaattctt
98041  ccttcttcca catattattt tcatcctaca gttttattgt tgctaaagta gtttcagcct
98101  caaaatRtat cagaaaagga ccacccagtt atatatactt ctattcatct gagatgggac
98161  aagctctttg gtaactgaaa tttgtcagat aggcccaact tattttcgtt tttcttgctt
98221  ttttgtacca tttctccctc ttttaaattc tacttatgtt ttgggattca ttcaagtaca
98281  ctactttcaa gataacttgt
```

FIGURE 3-A

>14:71227101-71317000

```
1       aaacccagag atgttaaaag acacattaac cagttgcaat gtagggttcc catttggatc
61      ctgatttaaa ctacaggggg aaaaatgaat gttggccaaa tatttgatga tattaaagtc
121     ttattgctaa ttatttgagc tctgatagtg attgtgagtK tgtttgtttg tttaagttct
181     ttaccattta gagatacata acaaaatatt tacagatgaa atcacgtgct tttgtggatt
241     tgcttcaaaa taaaatgagg tcagggagaa aagtaaggag tgcgtataga tgaaacacga
301     ttgaccatga gttgaacatt gttgaagcag ggctacatgg ggatgtgttt tactattgtc
361     ttcagtttta tatgaaggtg aaatttttcc aaatgaaaca tttttaaaaa catatccagg
421     cagagccaga aaaatcacat tcctgtagct gtctccctaa taaaaaggtc tttattctgt
481     tctttgcagc ctagaagact agtgtcactg tagctgtagt aataacagta atatcaataa
541     taataggcca ggtgcagtgg ctcacgcctg taatcccagc actatgggag gcccaggcag
601     gcagatcatg agatcaggag ttcgagacag cctgaccaac atggtgaaac cccgtctcta
661     ctaaaaatac aaaaactagc tgggtgtggt ggcgcacgcc tgtaatccca gcgactcagg
721     aggctgaagc atgagaatca cttgaacccg ggaggcggtg agtcgagatc gcaccactgc
781     actccagcct gggcaacaca gcaagagtcc atctcaaaaa aaaaaaaata acaataataa
841     taataatacc cagcattttg gaagcacata cttgctatgt gctaaactca tttaagagtt
901     tttataaatt ttctcattta attcagacga aatcttatgt ggtaagtact attattattc
961     acattttaca gatgaggaac ctgagggaca gaagtttggt gacaaaggta agcctcacac
1021    ccagatctaa tggcagattt ccggcttgta accaacaccc aaaggtaaaa gatgaagatt
1081    gctgtcatgt cctcagtctt gccttctttg ggataaacac ctttagctct ttcactcccc
1141    accccacccc acccctgcca catggcttca atgccccaca ccatccttgt cacattcctt
1201    gaaacacctg ccaggagcca taaaagtagc tgagcttctt taaggtatct cagaggccca
1261    tgtgcagctt gccccacccc gatcctaccc aggctgcctg gtttcccaga caagctgcgg
1321    tcctcctccc tccaggctgg gagcctgcac atgcgcatgt cccaggagtg gcgcagctca
1381    cctgggcctc ggtgcctctt ctccatccag atgtagcagt tgttctgggc caccccagtc
1441    tgtgagtcca ggaagggaag acgcacgctg cgctctgcac acagccgtga gttgtaactc
1501    cggcagtgct caatggcttc cttgtagaac tggtccccga gcctgccaga gtcagagagt
1561    gaaggggtga ggccagggaa gactgaaaca cacccagagg gaaacacacc cagagggaaa
1621    cacacctccc tggaaagact gaagacctcc cagttatgaa gaccaaggca atgagcacca
1681    agagccttgc caatgagcag aatgcaactc aatgtcaggt cttgactaga ttttatttc
1741    tatcccaaat ccaccaggca acttagaaac tcctttaaca aagagcatgg gcttccaagg
1801    atgaatcacc agtaagagag agctgtccat gttctaggaa gcatgaggtc ccactccaga
1861    agggaagtag ctctggaaca accttctgca tggaatctag aggacatcac acttctagaa
1921    cctcctcact catattagct acattttatt ctcagaggtc accagcaaag gcactctagg
1981    tgaacccaaa tcaccagcaa agatactcag ttttccatta caacagaaga gaaacactcc
2041    atataaacaa aaactaggac agaaaaaaca ttctataaaa gtagggaaat tgcaggaaga
2101    agggggtgta gaggaacagg caccaatctt ttaaaaccat attacttaaa gaaaaataac
2161    tttccgatag aaaacaaaca gaattcaaca attttaaatt caaaataatt caactcatgg
2221    agtttaaaaa acaaatagaa agacctttct ggtttgctag actgagaatc tgattaagca
2281    cagtatattt tacgttcctg actctttggt atttttggac acgtggtgat gctactaatt
2341    ttaaggtagt gtccttttca aagaccaaaa atggagcatt gatataattg atcaaatgca
2401    aagcactaaa ataatcaatg agatcaaata cctaccaaat aaaacatagg cacaggttaa
2461    taaatgcctc aaagaaatac aaatgtaata acataactta tgccaggtaa attcaagcct
2521    cttactgggg tcgggggcca gtggacgcaa aggagtggag tgtgggaaga agaagtagga
2581    agggaaggag ccacagagtt cacaggggca aaagggagaa tcaagtgttt attctgaagc
2641    aacattatga tatctttttg ttttggttct atgtagattt cagaaagcca aaatgagagg
2701    ggaagtctct aaaaatatca tttcttaacc aaagtattta gaggaataaa aaaagcaaaa
2761    ttggggtgtt gggtttttttg ggtttttttt ttttttttg agatggagtc tcactccgtc
2821    gcccaggctg gagtgcaacg gtgcgatctc agctcactgc aacctccacc tcctgggttc
2881    aaacgattct cctgcctcag cctcctgagt agctgggatt ataggtacct gccaccatgc
2941    ccggctaatt ttttatattt ttagtagaga cagagtttca ctatattggt caggctggtc
3001    tcaaactcct gacctcaggc gatccacccg cctcagcctc ccaaagtgct gggattacag
3061    gcgtgagcca ccatgcctgg cctcaatttc ttaattgtgg taaaatacac ataacataaa
3121    atttcccatc ttaagtgtac agtttaagtg tacagttcag tggcattaag cacatgcata
3181    ctgttgtacg accaccacta ctatccattc ccagaactct tttcatcttg caaaactgaa
3241    acactatacc cattaaacaa taactccccg ttccgcctct gcccccatct ccagcaacta
3301    ccaatctgct ttctatctct atgattctga ctactccaag tacttcataa aagtggagtc
3361    atacagtatt tgtctttttg tgactggctt atttgactta gcataatatc ctcagggtta
3421    acctatatta tagcatatgt cagaatttcc ttccttttta agtcagaata atattttaag
3481    tcagaataat atttttaag tcagaataat attccattgt atgcatatac cacattttgc
3541    ttatccattc atctgtcagt ggatatctgg gttgcttcca cattttagct attctgaata
3601    atgctgctat gaacatgagt atacaaatat cttttcaaaa ccctgctttc aggctgggcg
3661    cagtggctca tgcctgtaat cccagcactt tgggaggctg aggtgggtgg atcgcctgag
```

FIGURE 3-B

```
3721  gtcaggagtt cgagaccagc ctggccaaca tggtgacacc ccgtctctac taaaaataca
3781  aaaattagct gggcctggtg gcagacgcct gtaatcccag ctactcaaga gtcgcttgaa
3841  ctcagaaggt tgtggtgagt tgggattgtg ccactgcact ccagcctggg cgacagggtg
3901  agactctgtc taaaaaaaaa aaaaaaaaaa aaaaaaaacc tgctttcagt tcttttaggt
3961  atatactcag gagtagaatt gctgagtcat atggtagctc tagttttcat tttttgagga
4021  accaccatac tgtttttcat ggcagctgta ccattttcca tttctaccaa cagtgtacaa
4081  gggttccaac tactccacat tctcatcacc acttgttaga aagcaaagtt ttaaatgcac
4141  ttttctatac tctttttaaa aagtcaacgg gtcaggttct aacaaagagc ttaaagcaaa
4201  tatcttactg tgggaaaaca caaccctgga agcttctgag gcatcatttg gaaattactt
4261  gttgacaagc atccaaaaaa aaacaaacaa caaaacaaaa acccttttgct ccttttttcac
4321  ctctgaacat cttcacattc ataagggagg cataatacgc tcagaaagtg tcctggcagt
4381  tgcaaaacca aactgtaatc aaattgtaat agtcacgcag ctggaaatga attggggagg
4441  ggagatgaaa gcccattaca atgctataaa actgaatggc acataagctg atgtaatcaa
4501  agctggtaac aaggatcatg agattgatca aaaataaaga accacttttt acaaagtgtt
4561  tgtgtcacac cagcttaaat gcttccttct gctatggcga tgtagctcag tggaacagaa
4621  ctttcacagg cagcttggcc cagacaagat gacccaggct gttcagggta tcctggatgc
4681  tccagccccc acccatgaaa cagcaagcag ccataccaca tcgatagaga actgacaaga
4741  acatttgaca gagaggtgtc caacgtgaga gcagcaatct gttttcccag gaccttctga
4801  cagactgcca aaaaaaaaag atgttaatta tcccaacctc agaaattctg ctgaaaagag
4861  aacccataaa atcaggcccc ccaaaaaaac aaaatgaaaa attgtatagc aaaactcatc
4921  tactggagat gatcactttc aaaataaatt tctggaaacc aatacaaaag aaatctacac
4981  agtaggattc caaattggac ttgtcctaat atgctagtaa ctcacttcca acagcgtcaa
5041  tccagacaag aggaacaaga aactgaactt gctaaaaatg caaagtgagg ttggccaaca
5101  ctggtgctcc acaccaactg gatcgagaat gtcaaacaaa gtggcacact gggcgaaaat
5161  taatttcatc ttaaatactt tctactttca taacatgcaa tgttattagt tacaagcatg
5221  ctgggggaaa atgtatattt ctaaatataa tctttgtgtg aatgatacct gtcttatgta
5281  aacagaaaag actttctgtg gactggaaca actgtattaa agatcgaagc tgcaaaaaca
5341  tgacattcaa cagcccctct accatcatct ccagggaagt ccttttaaac agaccaggtg
5401  ttggcagaga ttcggaacca aaatttaaag ccttttcaa gatataagag gagaaaagag
5461  atttcatcct ccctgataaa ttcaagaagc acttctccaa aagttaatta gccaaaatcg
5521  tcacagggct tgctatgctc tctagaagaa aatcaatgac tagttctccc taacccgtag
5581  aaagcaattt aaacagcaga ggacaaatgt cctttccaac atggtatttc tgccaaaaaa
5641  aaaaaaaaaa atcccacaat gatgggatcc caggagtgac atttacatta aagttgaagg
5701  cgtcagatgg agcagggaaa ggctggcagc aaatgactca tctgtggagt ggtgaccagt
5761  gtaccctggg ttggatcaca catgggacag tcccagccat cacctgcatg caccaaactg
5821  agggggctgg atccccaagg gatgattgat gtcttaggct agaggaagct ggcacaaggc
5881  agggtagaat gatgttctta tcaacaactg acttcaggct actcaaggct gcaatcagaa
5941  ataggaaagg caagggaaac aaattgctgg gctccccagg ctccccagac agcattctgg
6001  ttgcagcagt gctggctgcc tagttctagg gggctttgtg agtgatggag gtMcttctgt
6061  gatcccatct gaattgagaa gccagcccag ctctgagagc tgagtgtttc tacatgagat
6121  ttccagcaac cctccttgag cctatctgtg gagatggact ccaggggcgcc aaagaagact
6181  cagaatggga aagatgagac ttctccaaag tgactgccaa ccaggctcca aatgtgattt
6241  ttcagagccc atcaatcaca acccctctaa gaccctctgc tcctcttcct ccaacctcct
6301  gctctcaaga ccacacctgc ttgctccgat ttctcagatc actctcaggt ccagccagac
6361  acctgggctc ctccctttca ctgtctctct ggttctgtgc tctgacctga ggacaggtgt
6421  cctgccatac cagaaatgtg agtccaatag caagtttctc atcttctgcc agaagcctgc
6481  tcccgcccac tcaagtctca atcctcctag gggcgggggt gggggtgggg gccgaggcta
6541  tgctgccctc cctctcccta gcaacttggc cttgactctc cccttgcctt caaaccatct
6601  ggaactgagt tatttcattt atcaggaagt cgtcctttct ctccactccc cctgtgtggc
6661  cccaagtccc ctttgctgtg taaacaggac cttggcccat cactactatc tgtgcacaga
6721  agcagagata agcagagctg aagagaagcc ctggcctgtg gcaggattgt gggggctctc
6781  cttcccccag aggcgcctct tcctcactgt gggcaccaag ttttcctgct gtctcactca
6841  gctgttcaca ccgccagagc cgcccacaaa cacacacgaa agctcatcta caccaatatg
6901  ctaacttcct gctatttgca tttaaacatg tttaactcaa actgttcatg gaaagctccg
6961  tctcccattt agtcttttag aataactggc acacaggggg caatatctgg caaggtgcgg
7021  gaaagttttt attttccttg aagtggtcct gttcctaatg catctcttga aagaagacca
7081  agtttagtgt ttgttacaga ctgaattgtg tcgttcctcc tcacagaaaa aatccacatg
7141  ttgaaggcct aaccctcaat gtgatggtgt ttggagatgg agcatttggg aggtgatatc
7201  atttgatatt taccctgcc caaatctcat gttgaattgt aatccccaac tctgggggtg
7261  gggcctagtg ggaggtgttt agatcatgtg ggcacatccc tcatggcttg gtgctgtctt
7321  cgtgatagca agtgagttct tgcaagatct ggtcatttaa aagtgtatgg tgcctccctg
7381  ctgactctct ctctcttgct cctgttttta tcatgtgaca tgcctcctcc ccctttgcct
7441  tctgccatga ttgtaagttt cctgagggct ccccagaagc caagcagatg ccagcactgt
7501  gcttcctgta cagcctgcag aacgatgagc caattaaacc tgttctttat aaattaccaa
```

FIGURE 3-C

7561 atctcaggta tttctttata gcaatacaag aatggcctaa tccaggaggt aatcaggtca
7621 ggaggatgga gccctcatga tgggattagg gcccatataa ggcagagaca tgagcaagct
7681 ctctcactct ctcaatctct gccatgagag gacacaaaaa ggcagctgta tgtgagctat
7741 aaggaggact cttacctgga atctgaccct gctagcaccc tcatcttgaa cttcccagcc
7801 tccagaactg tgacagacaa ttatttaagc caggggtccc aaacccctgg gccacagagc
7861 agtaccagtc catgcctgtt aggaactggg tggcacagca ggaggtgagc tgtggggtac
7921 aagtgagcat tgctgcctga gctctgcctc ctgtcagatc aatggcagca ctagagtctc
7981 atagaagcgc aaaccctact gtgaactgca catatgaggg atctaggttg cccactcctt
8041 atgagaatct aatgcttgat gatctgtcac tgtctcccat caaccccaga tgagactgtc
8101 tagttgcagg aaaacaagct cagagctccc actgattcta cattatagtg agttgtaaaa
8161 ttatttcatt acatattaca atgtaataat aacagaaata aggtgcacaa taaatgtaat
8221 gcacttgaat catcccaaaa ccatcccctc cccgacccca gtcggtggaa aaatgatctt
8281 ccatgaaacc agtccctggt gccagaaagg ttggggactt ccggtttaag ctacccagtc
8341 tatgccattt taatacagca gctggacatg actgaggaag tgtttctgga tgtttcccaa
8401 aaaagggcta aggaggacag gtgagagaga ttccaatccc ctttccttcc acatttttc
8461 ttcttctaac aaagacaaca ataataacac tataatatca atttcaactc aattcaatgt
8521 ttagtatttg tcttctctac tagactataa actccatgag ggcagaggcc acatctgcag
8581 agccactgca cacagacact caggctgaac aactccagca agtaccattc acatcagcgt
8641 caatatgaag gccacccccct ggagtggtgg agcccagcac tcctgcatac ctgcctggcc
8701 ttcttgctga tttacccgca gtgttcagca cagagcctgg cacataacag gtgcccagta
8761 agtctccatt gaaaagcatt tgacaactca cattccactc acctagggca ggcatttaaa
8821 ttacagcagg aaggctaatg tggcatgctt gtgcacagtc tgcagagtct gtgactggca
8881 atttaaactg gatggctgtc cacttgcagc ttcagaaaaa cattgagaga atgtctgcat
8941 tcactctcct gtagtcctca tcacccagca cctgggacat tgcagcactg ggtggcctaa
9001 ctgcaagcgc tatagcccta ctcttaccca tcctgacaaa aggaattctc cctgaaactc
9061 agtgtttagt ctttcaaact cctcagtgta gaacccaaga tggctcccaa tggtgcccag
9121 gttcctctgc tgggctgggg atgcacccgg aaaaactatg tattggggta aagtggaaaa
9181 tcaagacttg gaattggtca tgtctgagtc cagctccacc tcttaatagc tacacaaccc
9241 tcagcaagtt acttaccctc cctgggccta aattctacta ctcacaatat gggtaatata
9301 atctctactt cagacctgag aaaatttaat aagctagaac ctctggcaca cagacataaa
9361 agtctgcaat caatagccat tcctttctcc ttccccaaag tgcagtactt gctgccactt
9421 tgtggcatga ccctctgcct gagcaggctg atacccacca tgtgatgcaa cactcactat
9481 tgcggtgtcc ttgaggctgt cggggaccca ctcaaacctc tgtcccagga cagaaggcct
9541 ctgggtcacc aaggctgagc ctcaggcatc caggagagca gggcagctga tgcatggtga
9601 ccccagcaat ccacccaagg actgccctaa gcacagagac cacatctggc ttgtttactg
9661 ctctctccca aagcctagta catgctgggt catgcctgcc acattgggta aatgatgaRt
9721 gaatgagtga atgaatgaat cctcctctga ccccccctgca tatatgaaac agacagccct
9781 acccctgaat ctcaacaaaa cactttactt atatccttcc tcccacctgg aatttttgta
9841 ttctgtcctt gtcctggctg atcttagcct agtcaagagc tggctgaaaa cttgcttctt
9901 ccaggaaggt tcacaggtta tctctaccag tagggtttgc tctctggctt ctctgagcta
9961 ccttatctgt cttactttat ttttgatggt ctttgccacc ttccttatct ctgctcactt
10021 aattagattg gaagtatctt aaagacagga actgagagct atgtgcttgc atgctaccca
10081 accatgcaca gtactgggaa tgtagaagca tgtgatagat acattgcttg actgaactat
10141 catgtcatcc tcagcaaact aatgaaagct tcacagttga gaggtgcttg gaacgttcta
10201 actttctcca tgggaacatg aggtgagcct gaagttatct ggctacttct ccagagccca
10261 ggattaggca tctgaaaata gatatcctcc aagttcaaat cctattgagt tttttcccat
10321 atccttaaaa aaaggtcact tcagctcctg agattttcat ttctccatgt gcaaattcat
10381 gctaccctgc aaggtgccac agaactgaca ttctccttct ctgaacttta caagaagctc
10441 actgattctc aatgctgctg acatcagacc tatgccagga Raaccagctc cttctctgca
10501 tgggtagaaa tgcccttctt gtgcttagtg agaatgccat caccacatgc tccaaaagaa
10561 tgctctcaaa ctcaggattg tgaaagaatc gctaaggggt gggaagctgt gcaattctga
10621 ttctagaggc ttgggatgtg atccagacat cagctcccca gtgatcctga gaaatWtgag
10681 actagctgat ctaaagatag ctccccaacc tctaaaacaa gggcatgaat gagaggggac
10741 ttgtgttgct cctcagccac catgtttgaa tggcagagct ggacggccca gttagcagac
10801 tttgtttcat tcagctggtt ttgtttctgt gggaaaactt ctttcaggca gctggcatgg
10861 attaccctgg cagcactaac ctgcagcaga ctaatcccac aaaagtcctc atccctaaca
10921 aaattgccaa aagccctggt caaggcaaat gtatgaagat ctgacatcct tagtcattct
10981 gtgggtggct ctgaccttca aaacatctca ggaagggcac attggggaca aaggtatccc
11041 ttgggaaagt tatgtccgac aaacagcttg aaccctaaaa catccatcat gacttcgatg
11101 tggctcagtc tgatgttcag gcccagtaga cccaataaat ccctttcca gcatctaccc
11161 tggtgaagac ttttcaggga cttattatgc aaagctgcct aagcaagaca gaatttatat
11221 tgaacctggc catgatttct gagccgccac tgggtgttct gatgacatac acatctcgag
11281 gttgaagtca aggagctgcc aaggacaaga gactggtggc aggataaaag cagtgtgatt
11341 ccttccagta aataatggtg gggtggcggg ggaggtggga aggggacgct ccggggagga

FIGURE 3-D

```
11401  agaggaccgt agcttttctt gtacccaagg tttgcctttc aaatggcatc cagacgtgac
11461  caaataaatc tctcctctgg aggctaaaga atgtgtacaa tgcaacaaaa gtgaggtggt
11521  ggagaagcct gctttatcta gggtgaatca tacctagaaa gaaaacatgc atccgtcaca
11581  gaaatttcat tgattgagaa aatattttta gtttcctggt ccacagactg agtccttcct
11641  gaaactctca ctatgagctg ggcactcgga tgttgcctcc tgagtgtgat cgcctataga
11701  agcagactat tgattagaga aaattcttag acaaacagat ccccagccct catctttgcc
11761  actggaatgc aaagggtgct gccctcccca gtaaagggaa tatcggtctg gtatttaaca
11821  cggccacaaa gaacatgttg ttctatagtc aattctctat acagcagccc aagtgatttt
11881  ttttttttt tttgaaacaa ggttttgctc tgttgcccag gctggattgc agtggtgcag
11941  tggaggagtg tagcctccac ctcccagatt caagcaatcc tcccacctca gcctccagag
12001  tagctgggac tacagacaca cgccaccaca cttggctaat ttttgtattt ttttatagag
12061  gcagggtttt gccatgttgc ccagcctggt ctggaactcc ggggctcaag cagtcagccc
12121  gtctcagtct cccaaagacc tgggattaca ggtgtgatcc actataacca gccccagtga
12181  tcttttaaaa acagaaatca gggctgggca cggtggccca cgcctgtaat cccagcactt
12241  tgggaggctg aggtgggtgg gtcacgaggt caggagttcg agatcagcct ggccaacaca
12301  gtgaaatccc gtctctacta aaaatacaaa aattagccag gtgtaatggc ggacacctgc
12361  agtcccagct acatgggagg ctgaggcggg agaatcactt gaacctggaa gacagaggtt
12421  gcagtgagcc gagattgcac cactgcactc cagcctgggc gacagagcga gactccatct
12481  caaagaaaaa aaaaaagtgg aaatcaggcc atggcattct ctcctactaa cacctccaat
12541  gtcttctcac tgcactcaga gcaaaatcca aactccttcc catggctaca ggtccccagg
12601  ctggctgggc tcacctttcc tgcttcatcc tgtcccctgc ctccttcccc gctgggctcc
12661  atcctcactg cactcctttc tgttcctctt ctatactaag ggtatctcac ctcaggacct
12721  ttgcatatgc catctcctct acctggcatg acatgcttct cccagatctt cttgtagctg
12781  actgctcctc tgagatgtct tcttaggttc accctatctg gaagtagccc accccaccc
12841  aaccctgccc taccctttgc tatctcatat actttatttt tttcataaca ccactatcaa
12901  agtttttctc tttgcctact gatcatctat acacaactaa attgcaagtg ccacgagagc
12961  agcgtccttg tctcatcaga ttcccagccc ccaacgcaag acttggcaca taggaaatgc
13021  tcgacacata tcactcaaca aaaaaatgaa agctctaggc ccacctgcaa ctgacagctt
13081  gaatgatacc gaatcctggc cttgtagaag gctggaagtc agtgctggca ggagaccccc
13141  agcacaggga ataagaatca cacagctgcc accacctgat tcctggccct ttctgcccaa
13201  cacaacataa atgacagcag gtcagtggca cagtccacag tccacactca aatcttgatc
13261  atggccaccc acatacccca aacctggaga gagactgggg tgaggtagat aaacgatgcc
13321  attcaagaca tcagacattg ctttctatgc ccaaaaggca cattccaggc ctgatattgg
13381  caccggaaac tagcagtgcc agaagatggt ctccaacata cccatcccat gtccctcaaa
13441  gggcctctct ctggactaac ctcttctgca tggtctttcc tgcagagctg atccgtgtcc
13501  tcctgcagag ccctgggccc acgggtggtg ccaggggcag cggcaggatg ggggtggggc
13561  tgcttgacag gaaatcaatg ggaactgaag gtcatccgga gggcgcagtc tccccatcac
13621  gcacctgctg gcatgggatt tgtggtcctg acctcgcttg ggggcctctg ccctcctgtc
13681  agagttggtt aaaatcagcc tgcagaagca aatgtcctgc acaccaaatg atctccaggc
13741  agcccttcta agggacatcc agccaaacaa agaagcactg agtctcattt cacccgccgc
13801  aaagttgaga gtcaccccaa tctctcttgg cagcctccta ggaaaactgt ctcagacctg
13861  ccatttttgg aggcacatac ttgtaaatag aacaaagcca gactgaccaa gaggctggtt
13921  atgtccccaa acttaagttg cctggtggtg gaggaagtaa aaatacactg ggaaccaaga
13981  tgactgaaac tgaaatcttg agtgcaccct agaaatgcct aaatctgccc tcaacacctc
14041  tacatgccag gataacgaca aaggatacca aaacctctgt actgccggag ggaaatctca
14101  ttttcttctc atcccaaaaa atactgggga gcttggctga atctagaatc tgattttcct
14161  tcccccacta tacttatttt cccaggttcc cattctccct tgcttcccag acctttcccc
14221  tctaccatct tctttcccaa ttcgtccacc tcatcaagag tcatttatca tacaggccag
14281  gagcaaagcg gtcagcctaa gtctccaatt caaaagttga aaatgggcca ggcacagggg
14341  ctcacgcctg taatcctaac actttgggag gccaaggtgg gaagactgct tgagaccagg
14401  agttcaagca ccctggccaa cttggtgaaa aaccgtctct actaaaacta caaaaattag
14461  ccaggcttgg tagcaggtgc ctgtagttcc agctacttag gaggctgagg caggagaatg
14521  gcttgaaccc agaaggcgga ggttgcagtg agctgagatc gcaccattac actccagcct
14581  gggcaacaag agcaaaacac cgcctcaaaa aaaaaaaagt tgaaaatgaa tatggtatga
14641  aaataaagaa aaggccaggc acagcagctc atgcctggaa tcccagcact ttgggaggcc
14701  aaggcaggag aatcacttga ggccaggagt tgaataccag cctgggcaac atagtgagat
14761  ccccgtctct acaaaaaaat aaaaaataaa aaaatgtttt aatgaagaag agacaccctg
14821  gacatattta cagaaggata acaggatgct ggttatagta gcaatagtgg catagctaat
14881  atttattgag catttactat gtgtgccagg cactgttcta agcactttac atgtattatc
14941  cccatacaac ttgtaattgt attattccca tacaactcca tgattttata cccacccct
15001  cacacacccc actttacaga ttggcaaact gaggcacaga gagtcttacg tgatttgcta
15061  aggatctatg tggccttgtt aatacattac agtcttgcca gaaccttact cttggccact
15121  gcactgtcag cctcttacaa gggcgccatt acccccctcac aatgagtttt gcctccagaa
15181  ctagaataac ctgaacactt cctttcttgg aggcactaga acacttcctt tcttggaggt
```

FIGURE 3-E

15241 ttctatcagc atcaccaact tgctagctaa tcttagctat ttcttttctt tgtgcctcag
15301 tatcccaaac tgccaaatag cccagaaata aaccttctta tcacagagcc acctctcagt
15361 aactcaaaaa aagttaatat ggggcttcct gaaaatatgc catcagagaa atacaaaatg
15421 tcattatgca gtttagaaac caaatccatc ccaaatgtga gcaaggggcg atggacttga
15481 agaggccctg tggttctgaa catcaactaa ggcagaaaaa gagaaatgcc agaaagtcat
15541 aaactaaagg gtcagattaa agaatcccag aatgttagac actgtctagt ctggtgcctg
15601 gttttattgc aaaggaaatc aataaccagg ccaggtgcag tggctcacgc ctgtaatgcc
15661 aacattttgg gaagccgagg cgggtggatc acctgaggtc aggagttcga gactagcctg
15721 gccaacttgg tgaaactcca tttctactaa aaatacaaaa aattagccag gcatggtggt
15781 gcatgcctgt agtcccagct actcaggagg ctgaggcagg agaatcgctt aaacccggga
15841 agcggaggtt gcagtgagcc gagattgtgc cattgcactc caacctgggc aacagagcga
15901 gactccacct aaaaaaaaaa aagaaaaaga aaatcaataa ccacagaggt gaggtgacgt
15961 cctcgggaca taccgaggca gagctgggat cagatcctca gtccagactc ctagttcagg
16021 gattcctcca tcgtttatgg caagggtcaa caaactacag cccctcggcc aaacatagcc
16081 caggactaat ttttgtaaat aaagttttat gggaggcaag ccatgcccat ccgtttacat
16141 attgtctgtg gcagcttttg tgctgaaatg acagagtcga gtagttccaa cagagaccat
16201 gtggcctgca aagcctaaaa tatttaccat ctggaccttt acagaaaaag tttgccgact
16261 cctggttttt aacataagcc ttgaataatg aaaagcctcc accctcagaa tacaaaataa
16321 tagaatagga aacctaagag tcacacagct taagcagata gttttacaat tatctcagct
16381 tggtacaacc taataaaaac ccataaatac agttaataat aaggaaaaca agcactagct
16441 caccctacct atttaataat tattgttaac tacagcaagc aaaactcata gctgaaaatt
16501 cgtttgtttg ctttttttcta gcctcccttg aaagccctgt ttcatctagg aaatgctgtt
16561 ggccaatggg cgaaaggaca catcctagag gcgcacagcc ctaatttcct cctgttgagt
16621 tgcaggctgc tgctgaggca gctgaagact tgagagcatt gccctctctc cagagcggac
16681 tcagagagag ttgatttttct ctgtcaaagg cgatggcgcc aaaaccctcc tgtgcccctc
16741 actccagata gaaccctcag gaacagccag gcatggaaga gactgtcttt atggcttcct
16801 ctccctactg ggcctccatg ccctgcttac acccctttcc tccccagaaa gtgatcagga
16861 tatgcacacg gcccacagaa gcccagccca cagctatgcc atcattctat catcccatgg
16921 aatagggaaa agcgggtctg cccctgttag gctgatgtgg gaatcaagag acctggattc
16981 cagattggag ttggaactta gtcactttgg gtttttggca aatctctcaa gccctctggg
17041 cccaggtctc ccacgtcaaa tggagagatc gtcttcacat attccaaatg atggaacacg
17101 ggcagagcca aattaagggg tggggcaggc caaccacgtg gcacccgggc tccagtctct
17161 aatcacacta agtcattaSc aggaacatag aaaggtttac tagtgtcact caggcagaga
17221 actttgagaa ataggacctg gctttcatgg aatagtccag ttcgcatcaa ggtcccactg
17281 taatgattat tacacttttc cctctcaaaa gcgcccttgc ttggctgata aatgataaag
17341 tcaccctact aataagctgt ttgtggtggg agtgaggtga gacacgtggc ccaaggcacg
17401 gacagtaatg gcaatacatg ggcttctaag gaagacaagt tcagtttggt caagaggtgg
17461 ggtgcaggtc cggggccaga actgagctgc taaggacaag agaggagagg ctgcctctgg
17521 tggcccggct tctccatttg ccatgcacac agcctttgcc tcttctttta tttgtcaaat
17581 atttagagaa tattcgccaa agtgtgagct ccccaggagg cccacaggtg ctggagagaa
17641 gagagggctt tggagagtac tgtaaggtga aagaaaccta ccaacaggtg cctaactgcc
17701 tggccatgct aagcgcctga gccctgggac agggtatgat cccagccctt cacagtgcct
17761 gaccctgacg ggcatgcact tggtaagtgt tcattctggg gacctggaac aagcgaagct
17821 caagccctcc ttctgttctt ggaaaactcc tcctccccat ggcagctgca tctcgaccac
17881 caatcctact ggaccacagg atgcaggagg actacactgt gtcctggatc tcgtcattac
17941 aagagcagga ggtgagcact gagaaagatc cagaagaact catgcaagca gaggctgtgg
18001 tctgtgggga ggagggccgg gcagggcagc ggaccagagg aagaagcaac aaagaaggaa
18061 ggggaagagg aggaggagga ggaggggggc cgcagagtca ctgtgggtgg agggagagca
18121 ccgccagcac atgccccaca gccatcaata tctcttcgga ggacagatta tctggggaat
18181 ttatgctctc taataattac cttgcagagg gttccactta tcttaccaga ggaatctgcc
18241 agttgccaga cacacaggat tctagtcaat tccatccaca ctgccctccc cctctgccct
18301 cctccccacc tccctggctc ctgacttctg acctctgaat tctaaccttt ctttgcctct
18361 ccagtctagg gggaggttga tggtggaagg gtcacaacaa cttttaacag atgtaaaggc
18421 caacaaaggg gttggtgggt tgtcccttat cagtagatat gagagttaac gtcccaaagt
18481 tgaggccagg cctgaggaag tataggctct tgccaaagac agtgtgtgaa gggccaagca
18541 atgctttggg gcagactgac Wtgattattt ttgtcttcac ctgatgccat cagacctggt
18601 taggattaat tttatcttct agtcttcaca tccctcctca aagccttcct atccccaStc
18661 cagccctcca gtgaggcctc cctccaataa tggYgtggct acttattcga aaaccatcac
18721 gtgttgcttt gggttgctgt ggtttcttta aacgcacgtt gcagtcctgg ctgtgagctc
18781 ctggagaaca tgggtactgc atccgtactg aatcaggact acctgagcaa cagtaacaac
18841 atccagatgg ccctgacRgt gggcgggaca gtgccctccc tgcttagaaa caggaaccac
18901 atttcatctc cacaacaacc ctatgaaaga gatgctacta ctacccctca ttatacagct
18961 gaggacctgg gagctcagac agcttcagtc acaggcccac ggtcacttgg tcagtggcca
19021 agtcaggatt tgaaccaggt ggtcttgttt cagagtctgg gctcttaacc actccacacc

FIGURE 3-F

```
19081  acgctgcctc tgtaggagag gtttgctgac tgattcggac attttgtgtt gtatctatcc
19141  gcttatttaa gtcatccttt tctaaaataa taatttttt tgagacatgg tctcgctctg
19201  tcactcaggc tgtgcaggag ctctgtcact caggagtgca gtggcatggc tcactgcagc
19261  ctcgatcttc agagctcaag caattcacct acttagctac cccagtagct aggactacag
19321  gcatgtgcca ccattcccag ctaatttttt tttcttttt taagagagat cgtgtctcac
19381  tttgttgccc aggctggtct caaactcctg gactcaagtg atccacctgc ctcagcttcc
19441  caaagtgctg ctgggattac aggcatgagc ctgggtgcac ctgggctggt gctcctcctc
19501  cttaggactc cactcactgc tccctcctca gggggcctcc cccaaactcc ttagctctgg
19561  cactcattct ttttccatca ctctggcacg atgccaacta tcaaaaatta ttttataggc
19621  taggtgtggg tggctcatgc ctgtaatccc agcactttgg gaggccaagg caggcagatt
19681  gcttgagccc aggagttcaa gaccagcctg ggcaacatag tgagacctct tccctactta
19741  aaattaatta attggccagg cacaatggct catgcctgta atcctagcac tttgggaggc
19801  cgagatggga ggactgcttg agaccaggaa ctcgagacca gcctggtcaa caaagcgaga
19861  ccccatctct taaaaaataa taataaataa aaaataatta attaatgtat atgtttattt
19921  actacctgct tttctcaata gatataaaag cttcctggga gaaagactgg tctaatttgt
19981  tcaccccctct atcaccagtg cctaaagctg gtccctagca tatagtaggc actatataaa
20041  tatgagttgg atgaatgaat tgtctgatgt ccccactgag cattccataa tacatgatgg
20101  ctacttaatt ggacacttta ttttatgcaa actcatctat ctgttcttta tttattagac
20161  tttacagtgt accaggcacc ttgctgagta ggttaggatt tcactgtacc ctgatatgag
20221  gtagagaata ttgtctttat cattctcatg ttacagatga ggaaactgag actcagaggt
20281  tacacagctt gcccaagacc atacagaaat agggtaagac ttaaccccca ttctgtcagt
20341  ccccctgagt ccatgatatt aagcattctg cctcattgcc tctggctggg tcccctcagt
20401  aacatgattc tctttacccc tgaacctcta aaatgtcccc agccttactt ctccccccctc
20461  aagagcccca tcgcctcttt ctctggcctc tgtgtgctgg ctctcctcca ttggtccctc
20521  cagattctct ctctgccatt cttttccctg ctgggtgctg cggaagtctg tcctgtatgg
20581  tcagccattt gccctcttcc ctctgctcct ggtggtctgg gtcaatggca ggtgccgaga
20641  gagggaagag agacagggga tttcctctcc ccgccccttc ctgctgtgga gtgtctgcat
20701  ctgcccactg aaggtcacag ctcctgtggg ccagccctct tcatacagct acctctgcaa
20761  tctggtgacc ttccccactc ccccgaccta gccctaggct gctgcactag ccccccacca
20821  tgtgagtccc tgcccgcagt tttgaaacta gacatgttac taaacactcc tcagtttact
20881  tcgtgtggag ctgctgcttt gctggaaccc tgcctctcgg acatatgagt gctacatggc
20941  ctctgccccc cccccaggtc ccccaccact gccgctgccc cctggtcact gcagctgcct
21001  caccaggccc ttctttgctg ccagggtttg ttctctgtct ccaagaccta taaataaatg
21061  ttgatttgct ggcttgaagg tttgccctgg cagccccacc agggagagca gtgccctggg
21121  gagcccagca agcccacagc aagctgtcct cccaagacta ggagaaagca gagagcatcg
21181  ggtccaatgc aggggtcccc tatgctcgga gcttcccccc gtgcggggtg ttaagggacc
21241  cctcagatga tgtcctggtg cctcaaacaa agccgccatc acagcttcaa aatcacaata
21301  tcatttgttg agcaactact atgtggcagg ttctgggttg ccatttgcat acattcactt
21361  tcatgttaaa agacgagaaa aacaagttca cagaggttaa ggaatttaac caaaacctca
21421  aaactactaa gtggatgaac tggaattaga tcccaggtca gcccaaatct ataaggcctt
21481  aattccccac cagtacatgc tgcctcatga ccgatccctt tgaatgacag gaggcccaca
21541  gactcgaacc tagtccaggc aggcagctca gagatgccag ccacccatgg ctgggaggag
21601  atcaggtgga gcaggatgac agttaggata aagggcacag agggctccat gctgcctaga
21661  gtcactttgt ccaaggacca tggcacagtt tggccatagt gactgctgag cagccagggg
21721  gctggaacat gccccccacc cccgccccca accccgaccc actgctcctt cctgcagagc
21781  cagccaggag ccatcactcg tccttgggca cacaccagca tttcaagaaa acaagcctgc
21841  aagcacagac ccctctgagt caggcccagg actgtcacga ggagactctg aaggggaggg
21901  gagggacagg gtgacacaaa gaggggggaag gcctcctgcc gctggacctg gagccgactg
21961  tgcttcccta agtcctttct gcagccaaaa ggcacattcc ctgagaacac cgctcaagat
22021  ccaaaggctt tcgcttttta ttttttagga agcaatggca aaaaaagaaa agaaaaaaag
22081  aaagaaaaga aaagaaaaag aaatccacat ttctagtgtt gggagcgatt cttggaggca
22141  aacacactct cttgcacaca cacaccccac agcgcttccc ccacacagcc caaaatagtc
22201  cacaccatgg actcctggag aacccggctc agcacagcct cccgaccttg gtccacatga
22261  ctgctggcca gctctcagtc cccagggtaa caaagtactc caaggacccg cagtagcagc
22321  gcctgctcct acagaccatt tcacccctcc cagcgaagtc ccatctgctg tgtctcgatg
22381  gagccaggtc cagttagcag aggccacttc caatgaccca gtgtggccca aatctccaga
22441  gctgacctga acctctctcc ctcaacccct ccaaatctta gattgctgct gctgcttctc
22501  caggggaatg gctcccacaa cattttatag actgcaagga aaatatccag aactgagacc
22561  gttcttcaaa ataataattt aaaaaaaaaa aacactgctg aaacacagaa ctgggtgtta
22621  tcttaaaatc tctgcacaaa tagaatttga atgctctcac tccaatcaaa atcattctcc
22681  gtccctccct ttaaacactg aaagttcaat tccccctaaa gacaggattg caaacatgat
22741  taggtccaat taatttatgg agaaactcag aatgaaagga gagaaatcaa tagaactgcc
22801  ttgttctgat cactaccgca aatgaagtaa taatttgaag gagatctgga aagaaaggaa
22861  accaaaagtg tcacaccgca gccattggcc atttaacaag gcagcattgc cccatgtgga
```

FIGURE 3-G

```
22921  ctggctgggt ggagaagacc actcaagctt acaagacact cctggggaca gagggagggg
22981  agagacagct gaggtcactt gctttctctt cattgtacag catgaggctc actggtggta
23041  aattctcctc tttctcacaa aagcaagacg taaaagaaag taaaaggcaa aaagcatcca
23101  cacaagctgc cagtctaata aggacactgg taatctccag gcaggccaag gccatatggg
23161  ctccccctgc agcttcatgc taagtttatc tccatcctca cactctacct ccatctctct
23221  tccatatcca ggaaggaaaa aaggaaatta acaagccggc atccaacatc catcctgaat
23281  caccagctgg gccatgtaga gcatccaaag atggggtggg gaagccccag gggagagaga
23341  gacattgaca gatcaccgtc cccacctctc acttgcctgc agctcccttg catagatgat
23401  gctggcaaat tcaacgtgat tttcattgga cctgcccagc agtccagatg tgggagccta
23461  ataacaaata acaatgccag ctccatgtga tccttgcacc ttcaagagaa aatgccttgc
23521  tggccagctg ggttctctgg ccaccctcca gcccaccttg ctgcttctgg atggaggttc
23581  gtgaagctgc tcctggccta tccaaatggc cagtaaacat taacattcat ccagtcaggc
23641  aacatccatg gagtgcctat tccatgccat gctgccccat gtgggaggca gcaaagcgct
23701  tcagtgaaag gctggggcac tggcatgagg acaccaggtt tgagtcccag ctccttgctt
23761  acttgctgtc tgagctgttg aactctgtga gcttcactac cctcagctgc aaaatgggta
23821  tgataataag acccttctca tggcttcact gacaagaaca aatgaggtga cctcactatg
23881  ggtgatcagt aaatattagt ggttattacc acaatgtggg gtcctaccct ccactcattg
23941  ccaccaccag tgccagaaaa agccagctca tgagttaact cccaggcagt attatccact
24001  ccaggggcct ctctgtccac acagatgagc agaagcagaa tgcatgaaaa aataaaatag
24061  ctcaatgcca ttgctacaga ttgccccacc accaccatct caccagtcag cttcaagcca
24121  gcggggtcag cagtcagaag cacagccaag agtgtgggga tgattccgtg acaaaccatc
24181  cccagcagca gaaacctatc tcacagcaca catcctccca gaagtgggtc aagggcatcc
24241  attgcctttg tttaaaggca ctgggtcaca aaggaggagg agccttagca ggttggaact
24301  ttgggatatt ccacaggaca ggtagaaggg tgctgtgttg gttttctatt gctgctgaga
24361  caaattacta caaactttgt agtttatgaa aacacaaatg tacgatttta cggttctaga
24421  ggtcagaaat ccaaaacgag tcctaaggct aaaaatccag gcatcagcag ggctgtattc
24481  cttttggaag ctctaaggga gcagtttctt ccttgacttt tccagtttcc agcagctggt
24541  ggcctcttcc tcccatgtgt cacttctacc tccacttcca aYgtcatatc tccagctctg
24601  actctgaccg tcctgtctcc ctcttgcaag gacccttgtg attacactgg actcacccac
24661  ataagccagg ataatctcct caYctcaaag tccttaactt aatcacacct gcaaaatccc
24721  tcataccatg taaggtaaca tattcatagg ttccaaggat taggacRtgg acatctttgg
24781  gggtaaggtc acgcttctat ctaccaaagg cgtgttctgc tgctgctgca aggtacactt
24841  gaataccgtc aaagcaatgg tgaagaattg aaagaaacgc agttatgtcc taagcccagg
24901  aatattgaag ctgtcctctt tttgacaaaa aaaaaaaaaa aaaaatatat atatatatat
24961  atatataagg cagatggtgc ctaagagcat aggatttgga accacatcgc tctgggttca
25021  aatctaagct tcaccaccag tcccagctga agctattcaa gaccagaagc tcatctgtgt
25081  gaacttccat tttcccaaca ttgtagggtt gatgtatcct aaggaataaa gaacaaaatg
25141  cacgcagaat ctttcacacc atgtctggca cacagtgaat gcttagtaaa tggcaactaa
25201  gaaattgctc ggtactgtta ttctattagt ttggttctat cagacccagt cactaaataa
25261  accagtttga cgtgaagtac ttggctatca gcaaatacta ccaggccctg tcaagtccca
25321  ggaatcagag gggctgtttg gccaatgcag actgaccctt tcatgctgac gtcagggagc
25381  cccagcagag acaaccaaag atctagagac atgatcaaaa tttccctgtc aggacatggt
25441  ccagaaccat ggatctcagt ttactttccg cagaaagaag aagaaaaggt gaagcccgca
25501  cacttctctc ttccacaaac tcaggcccca ttcccttgta cgtctttctc ttataaacaa
25561  ttgcacttta ttttcaattg tgttgtgata aatttgagct atagcttaca gtcaacaatt
25621  tctatcacaa atctttctgc cttccaattc tgtgtatgta caaactgagg ctaggaccaa
25681  aggaacgaag actttatctt acatgaatca tttgattcca ttactgaaat ctctcaggtg
25741  aagttaaaat atatatacac tgggtctata ttagaatata tgagacttga atttgatctg
25801  ctaattagct gggtgacttt aagaaagtat gtaacttatc tgagcatccc tttccccatc
25861  tataaaagga ggcagaggcc aggtgcagtg gctcatgcct gtaattccag aaggcggagg
25921  caggcagatc acttgagctc aggagtttga gaccagcctg gacaacatgg cagaacccca
25981  tctctaccaa aaatacaaaa aattagccag gcatagtggc atgggcctgt ggtcccagct
26041  actcaggagg ctgaggtgga agaatcactt gagcccggga ggcggaggtt acaataagcc
26101  aagatcatgc cactgcactc cagcctgggt tacagagtga gatttattta tcactaaata
26161  aataaataaa taaatggaga caggaaaatc tgcctcttaa gactacagca tgggttaaat
26221  agcatgtaaa acctttggca tgttggattt ctagagcatg tactatgtct cagacactac
26281  aataaacatt gtatgtgcat taattcattg aatcctaata acaactctat gaggtgctgt
26341  ttattattcc tattttccta gaaagtagac ctcttggggc cagggatagt ttttttcatc
26401  ttcaactccc aactcaaata ccctgtgagt ggtaaatgca agctacaact atcattgtcg
26461  atagttgagt ctggaatcac cgcagacgat tttataaaga tggaccccaa gaggtaggga
26521  ggtttagaga aaaggaggaa ggagagtgtg caggtcttcc caatggggta aggcgaggca
26581  tgaggaagtg acaagaaatg ccaagccagt aagaagacag gacgaagcat cgcacctgga
26641  tagaagacta agtaaaacta ccttcctgtg gctaaggggt ctactaggat acactcattg
26701  acttgctttg gccaacagaa tacagtggaa gttagattac accagttcca agctcaggcc
```

FIGURE 3-H

26761 tcaagaggcc ttgcaagctt cccttctgct ctccgggaac cctgcctagt tgcttagcat
26821 atgaatacgc ttgggctagc ctgctggaga atgagcagac acagaacaca aacaagctat
26881 cctagatgaa gctattcatc taggacaagc tgactgcaga cacataggtt aaccagccca
26941 aatcagaaga accagctagc tgagcccaaa tcgaattgcc gacccacgga atagtaagca
27001 aaataaatgg ttgctatttg aagccattaa gctttatggt gttttggtat gctgcaaaac
27061 taacagaggc aggagtcaat gtctcattca tctttatatc ttcaacagcc aataaaagtc
27121 ctggcatgta atagttacac agtaaacgct tcatgaataa agtattgaac tgatatatca
27181 atacagctgt gaataaataa atgtcatatg ctggaaaact gaaatcatgt aagatcgcat
27241 aattctattg agatacttca gcttaaagaa tatttttaaa tggggttggc tgctacctaa
27301 ttttaaaatc cattttcttt tttcttccta gccaacatcc ttttccacaa agtttagtca
27361 aatagaaaag aaagtttcac atgaacatac aacgtaggat gYggtgatcg aaaactagac
27421 tggttgctta gagatccgag ttctagactc aactctccca ataactggga gaccttgcta
27481 attctcttaa ttaccatttc tacgtctata aaatagatat taaaaaatta cccttccctg
27541 tcaacctcac agggttgttg tagcaaataa gacagcatgg gaaggctctg agcacttttt
27601 accgtcaatg attctacttc cagccaggct cagtggctct tgcctgtaat cccagcactt
27661 tgggagtctg aggtgggtgg atcacctgag gtcgggagtt caagaccagc ctggctaaaa
27721 cggtgaaacc ctgtctccac caaaaataca aaagtagttg ggtgtggtgg caggcgcctg
27781 taatcccagc tactcaggag gctgaggcag gagaattgat tgaacccggg aggtggaggt
27841 tgcagtgagc caggatcgca ccacggcact ctagccaggg aagacagagc tagccttcat
27901 ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa aagattctac ttcctctata agcagaactt
27961 tacttttga tggtttctag tgtttttata atatatcctc ttaagagccc catttatcct
28021 tttgctgtct tccttttaat attctgtatc ccagtagcca tattttcaac atacatacat
28081 acatacatgt atgaaaccaa gataggaaat ggttaaaatt atccttgata gtggagaaaa
28141 ctattgctcM atccaaagag agctggtatt tctggcactg aggattacat gtgaccaagg
28201 tcagacatga tacctgcaag atgtatcctc tagctccaga gttcccatgg agaactcggg
28261 gtccctagtt ggctgcagca ggggctgcca gctgctaagg atgcagccag atccctttcc
28321 ctttccagag acttctgccc agtatgggaa gtcagaagct ccacactcca agtatggctc
28381 ttccacaacc tcagttcctg ccttgggcca agtcttgaga tcagtgtagc tgagctccga
28441 gcccaggtct ggcactgcta aatggaaatc cttcatcttg gcaactcctt ccaYaatggc
28501 agaaccaggt tggaagttca caaggtgctc caggcatatc agggcttact gccctctcct
28561 aaagctcccg aaacctttgg aagtcacagt tctggagtaa acatcaggtg tcctcatacc
28621 caagcatctc aggagtgtgc aggaggcagc agaaggtagc ctgtgtcctg aggcagtgca
28681 ttctccccat cccatgaatg atggggctgg acagaagaca tcagggacag caactcaaac
28741 cactgggcag gtggggctga atggatgggc aacactggcc gtgctcaggg ctctgggtct
28801 tccttgacag tgcaattgta atacctccat ccttccatca acccatcctc cctatggtgg
28861 acataggttg tctgcctgcc cagataccat tctacctcct tctagtaaca atacctccat
28921 tttctacctc tccctcacta tcagttgtgt aaaatcaacc ccactctacc ctccaacttc
28981 agagacaggc ttgtgactag gcctggccaa taagaaaacc ggccatgttc tagccagcca
29041 tggtgactca cacctataat cctagcactt tgggaggctg aggaaggagg attatttgag
29101 gccaggagtt caagaccagc ctgggtaaca tagtgagacc ttgtctctac aaaaattggc
29161 caggcatggt gatgcacacc tgtagtccca gctactctag aggctgaggt aagaggatcg
29221 cttgagccca gtgagccacg attgcaccac agcactccaa cctgggcaac agagcaagat
29281 catatatgta tatatatata tatgagaaag agacccacat cccttagtga tccgttcaag
29341 agtgatcagt tcaagggtga aaatgtggtc tgaggcaggc aaattagagt caatcctgga
29401 acttctgggg gaaagaggca ctttctttct cttagtatgg aaaggcagat aacaatatat
29461 acctgaagca gctgggagcc acccatcaag caggcctgtt tacaaaggat gccaatgtaa
29521 agaaaagtgg agtttagaga tgaaggagag atggagggag tcaagaggac cttgctgagc
29581 ccctagaccg agtcacgagt gaagaagata cactcacgag ctgttagtta catgagccca
29641 cacattcctc tgtgtgctta agcactatga gttgagtttc ttgccatttg caaccttaaa
29701 aagccctgcc tgatgcacat cccaccattt gatgttacca acagactcta ataaggctga
29761 ggaagaaccc aggggcccca tccaaatacc tctaggctgc ttgaactcaa atgcttttga
29821 aataaggatg gttcactgtg gctgtcagat gccacacatc aaacctgaaa ggaaattcta
29881 gccagggaat gctacagatg gcagctcacc acggccccca ccactcctta ttgtctttgt
29941 caagcccatt tcctcattta ctttacctgc ctggtccctt taaatacata agtttaccat
30001 cccttattag atggataaat agataaatgg atggatggat ggatggatgg atggatggat
30061 ggatggatgg acggacacat ggatggatgc atggatgcat ggatggatgg atggacacat
30121 ggatgcatgg atgcttggat ggatggatgg atgcatggat ggatggatgg ttgcatggat
30181 ggatggatgc acagatggat gtatggatgc atggatgcat gcatgcatgg atggatggat
30241 gtgtgcatga atgaatgcat gcatgcacag atggatgcat ggatgcacag atggatgcat
30301 ggatggatgc atggatgaac ggatgcatgg atggatgcat gaatggacgg atgcatggag
30361 gaatggatgg atggatagat gcatgggtgg atgcatggat gggtgcatgg gtggatgcat
30421 gagtggatgc atgggtggat gcagggatgg atgcatagat ggataggcag acaagcaagc
30481 agttatgtag ttgcaagttg aacctacaca caaataggca tgatacaact aaagaattaa
30541 aacggcagcc cactgcagtc agggggccat ctgacaatat aggaaaggag ataactatag

FIGURE 3-I

```
30601  gcttacttct gtatggtttt gggtgtctgg aataacggga gatgactctc tgggtagcac
30661  tcaccctgta gttgtgctcc tccctctcaa gccacagtgg ttaacaccac cttaccaata
30721  gtgcagtaca aagagaagcc agtcctggca ttcacaggta gctgtgaaaa gctggcccta
30781  aaaggcagtc tcgctcaatc cccagcatct acttggatct gggcaggcag ccacagaaac
30841  agaagaagct acagaggatg caggcactgg gttttggtgt ctggacatgg aagaattgca
30901  ggccaatatc cacaaatgta tccgtgaggc tctggtcttg aaacaaggat gtgaggaccc
30961  caggatggct ggtgagggac actaagtggc tcaaagaact gagaaaaaaa taaaaacata
31021  gcagcaatag agaccagagt ccgagccagg acaaggcctg tccctcagcc acctggctga
31081  aggcatcatc atgacctcaa ggaggtaact ggtgagtcat acagaaccac actgtgcact
31141  cttcaccatg gtttatatgc tcacagttca ggcaattgct tagaaaactg catatcgttc
31201  atgcccctat gaatcacttt cctccttttg ttgtctttgt gaagtcatca caaagttctg
31261  agtggaagaa ttcttccagg gactgccaag gagtgacact gtagccaggt gggttcagct
31321  ctcactttgc tcaggatatg ggtgtccagt ggcgacagat caggaaaata gccccatact
31381  aaatggtgcc tggacaaggc ccatacacac ttggatatgc tgtcctgaac acaaggcaca
31441  agacacaaca cgggaggacc tttcatctca tgtctcttac agccagggac tgaaggtgta
31501  ggacagtagg agaaaataga gtttcacata aaacacccca aggacatcaa agaaggctag
31561  ctgccaagag gtaggaagaa ctagattctt agggacttat ccttggggat tgataccacc
31621  cctgccaggc atccaccttt ggggtgactc agagccaaga aaacaaaggc atgggataag
31681  ccatggaaaa caaattcctg ctgagtagca aagagagtgg caccaacaag tggcaccatg
31741  tgagtgacaa gagcaagcaa atgctcacac agggacagct gggtgtgcag ggccatcaca
31801  gctcccccta cccccaccac ctgaacacta cgagacaaat gggactcctg gtttctgccc
31861  tcaacatgct tatatttgag agaggccaca ggaccagcga gtgcagagtc atggcgaact
31921  gtacacaacc tggagtcagg agacccatgt tctcaccctg ttctgttgcc tcgggggccc
31981  tgaagcccat cgctaagtgt ttKgggctcc agtttgctca tctgtcaaga gagaaccctt
32041  cacacagctc ctaagtgctg tgagtccctg tctctcagac tggctaatct ccaagggcct
32101  gccaggtccc acatgcatgc tcttgagagt cacatagata tcaaagctca aaattagctc
32161  ctgggagtgg gagcaaagct tctgagcccc cagcccctcc tgtgagcaca tacctcaagg
32221  cacatgcccc caacagccca gaactgacac atttttgata gttcctgagt acatggattt
32281  ccagccctgg aaatgtttgt tttaacttac caggagaaaa gaatttggat tctttttaa
32341  atatacacaa agggaataca tttgtatatt ttcagataca aagtggatgg gaagcaaggc
32401  taaagactgt tcttttccct tagctcatgt gagtgctatt cctcacggtg aattctccaa
32461  agaccatgaa tggcagagat ctttgatctc ccttctagcc tgtctcgatg tgatagcctg
32521  gacatgggca gctatcacgc cacccttcag aacctcagca cttcaatgac cacatcctgc
32581  catcatcacc cccatggagg tccagggccc ctgccccact gcctacctct ccacctgtgc
32641  ccacctctcc acctacccgc cccttgtcac attgacagat aaaatacagg actcccagtt
32701  aaatgcaagt gtcagatgga gaacaaataa tattttagta taacatgtgc cttcctaggc
32761  agcacaactg aagcaggccc cttccaagga tagaaaagtc ctcctctctt cctaagtttg
32821  tgtccataga atctcatcca gtggcctttc tgtcccctgg cttcaaatga aacctgccac
32881  tgcctctctc tcagggtgaa gtgtgaagac cgtggcttcc aaggccaata ggcatggagg
32941  tagtcatccc attcagacgc caatggttct tagtcaaaaa agctgcttat ggctggagtc
33001  aaaaacatgt ctaagcgtga ctatcagccc cagtgcagaa gagtctgggg agagggagaa
33061  acaagaaaca gatattacca caaaacactc caagggaaat ctcactacaa aaaaagaggt
33121  tgttgcctgg tgatggctgg caatgggcca tgaaaaccta gttctgattt ttagatacaa
33181  gacacaagcc caacgtgtca cccagcctgg atcccataga ggtcacatgg acaaaacagt
33241  acttgaagat ctttgacctg cctaggtggg caactggagt ggaagctcca caagggtatg
33301  aatgctgtct tttgttttcc tcatctctat ggactagtcc atctagtcca ggactagcct
33361  cagagtgagc atccgatgat cgtaggaaga ctgaatgaat gcgtgaacac atgaatgaat
33421  gcctacattc tccaatatgt cccagaaagt gtcaagagct ttgtaacgca aaaggctctg
33481  aacaactcaa ataagaaaca caaacatgtg cagtgaaagg gctaggcctg gatggcctgt
33541  gaggtcattg gcaatgtggt gagcccttga gcctgtcata cacaggcaaa taacttcact
33601  gaaagaattt taaacgctac ctccaacacc ctagctcagc atctatggtt gaagccacca
33661  tgcccagtca tggtgataac cccagaaggg aagggtgcag agacgccaag aaacaagaac
33721  ttggagagga caatgaggac caaatgactg gggaaataag ggatgagtgg gtaggaatcc
33781  aaagaacaga ataggaaaaa taactcaaga tccagcactg ctttccagga cacacacaca
33841  cacacacaca cacacacaca cacacacaca cacgcagaca aagattcccc ccttccagat
33901  ccacactgcg tgcctcctgc cagccccatc ccatcatcct gtgacagtgc ttctttgtgg
33961  aatagaacaa taggtacatt atgccaacct ataattatcc taggagaagc gatggcgaag
34021  tcaaggcatt atgtgaacac tggaatactg gataccaccc caccaggagc aaacaacaac
34081  aaaaacaaca acaatccaca aaatctctaa cgcagataaa gactgaaagc tgcctaagcc
34141  agtgcccttt cacattcata tgagggtggg atttttcct tttcacatga ttatttaagc
34201  acaaggcctt ctttcatgga attgatagag tattgggggg aggggctgag aggttgggtt
34261  tgtttgtgtt tttgtttttt ttcccaataa ggaggcagaa agtagaggag gtcgtgagca
34321  cgagaaagtg ccactgctag ggtatgacac atagagccct caaaggtagt tgaatctgat
34381  tcaactgagt gaaatcattt cattaataga gattccttct cactactcct gtgaaactac
```

FIGURE 3-J

34441 tttccaggga cgttggtata aaataggctg agtcctggga ggtgtcatgg aagacagaga
34501 caagaaatca gaatatctga attcttatat cacagccact aactagctgt ttccttgatt
34561 tattcaagag gtatgtgttc cctatctgcc acatgctagg catcaagcta gcaacagaag
34621 gaaaaaacaa agcagtagtc actttcctct cggagcttca gcttcctact gtgtaaaacg
34681 agggagtccc acagccctgg acacacacac acacacaaac acacagacac acacacactt
34741 ctaaatacaa ggtgcagccc cagcctccca caggacctgc ctatacaggc catgactgaa
34801 gaaggggaa acctgcagct tctaaaactg cacagaaggc tgggcacagt ggctcacacc
34861 tgtaatccca gcactttggg agggcgaggt gggcagatca cctgaggtcg ggagtttgag
34921 accagcctga ccaacatgga gaaaccctgt ctctactaaa aatacaaaat tagccgggca
34981 tggtggcagg taactgtaat cccagctact caggaggctg aggtaggaga atcacttgaa
35041 cccaggaagc agaggttgcg gtgatctgag atcacgccat tgcactccag cctgggcaac
35101 aagagtgaaa ctctacctcc aaaaaaaaaa aaactgcatg gaagcccaaa cccccagggc
35161 aaatggaaca gtcacagcat gccatgcacg cacagggtca gagcaagagt atggccaggc
35221 gcggtggctc acacctgtaa ttccagcact ttgggaggcc gaggtgggca gatcacttca
35281 ggtcaggagt tcaagaccag cctggccaac aaggtgaaat cccgtctcta ctaaaaatac
35341 aaaaattagt tggacattgt aatgggcgcc tgtaatctca gctacttggg aggcctgggt
35401 attagaatcg cttgaacctg ggaggcggag gttgcagtga gttgagattg tgccactgca
35461 ctctacctgg gtgacagagc aagactctgt ctcaaaaata aataaataaa ataacaaagg
35521 tgtgaaaagt aggatgggtg tttcctagca catctctagt ttctgagggc tctctctaat
35581 ggggaccYtc ctcttgaatt gtggtcccat gggttccaYc tcttctggga tccagtcagg
35641 gcttcttcct tcattctcag gcaccttcag ccttgaaact ccttggcctc ctttcccttc
35701 accttcctgg tgaaacataa cagagggtct tcgagacaat tcctttaaat catcaaaggg
35761 cctatctagc aatgtctaac tccattcctt aacgtttact gaaaagctcc tcaaatacat
35821 gatcctcccc tctacccgtt cccctcagct cttagYcatc cttcctcaga tgctacttt
35881 tcatagggac taatttcttc cttcttgcca aatccagtgg ccacccgtgg gtggtcctgc
35941 tgctcaacca acgtatgaca gtattagcca tgcgctcctt gaaacactca gagccttgat
36001 tctactcatt catttagttc ctccccaaaa gttccctgag cccctgctag gcagtgggga
36061 tttggcaata agacccagtc cttagtcctc tgctcatctc tccacactcc gctcttatgg
36121 ctccaactct cacatcagca aagggctttc aaagactcaa agcacctcca cactcattac
36181 ctcttgtaat aacaaaattc aactctgcaa actttgcaaa acactttata tcgtaatcct
36241 tttttcccca cccScactgg gcgaaagaac ctattcttgc aaccaacact atcctgaaaa
36301 tgtggaaaac ctctaatgga agtgcaacca agtgctcctg gaattcgaca tgcttttta
36361 tgtccttccc acccctagtt ctcccctccc tgcccactct cccttcgcac ttcgcttgcc
36421 taccttacag cagagttcat tttgtcggtc tcttctaagt ggtgggctcc tggagggctt
36481 ccaacggact gatctctttg tagactccat attactcggt gcttaaatat ttgttgaaat
36541 gcagcaagat cagggagtaa ttcagcacca attgtttacc tgatcctgga aacctcatct
36601 ctgccccctc cattccccag gagagtctgc ccatttgaaa actgttcctg ggtttatttc
36661 tagcatggcc ctggagaggg ctgccctacc ccgcagggtc agggaggtga gttttctctc
36721 accttgggta ctgcagagcc caccaccagt ccaccctgaa acctcaaagc aggcccacag
36781 ggtagatatc tagggacctt gaagcaaaca tttcacctat tcaggacaca aactaaataa
36841 gcttgtgtcg cttagtaatt ctgtaatgtc ttacttatgc ccttgggcaa actccttaaa
36901 atatttgttt ctgttttccc atcaataaaa tgagaattga agcacctatt tcttaggatt
36961 attgtgaaga tttgatgaat taacactagg aaatgtgcct agcactgttt aatatgtggc
37021 agaggctcaa taaatacatt attattatta ttactattat tgacaggtta aaggcctgtg
37081 atgatactga ctccagcata actctgatac tgccccttg actcggctag ttttgacaga
37141 gggccctttg acaacaccta aaaaacaaac cactaaactt tcagttttta aaaaaatacc
37201 aggtaagctg ggtacagttg gctcacgcct gtaatcccag aacttgggga ggccaaggct
37261 agaggatgat aggaggatct cttgaggcca gtttttaaca accaactctc cagYgaacta
37321 atccattcct gccagaacaa atccagtctt acaagagtat gaatattaga gctacaaggt
37381 accgccaaga gacagggttt tgctttgttg accaagacca gcctggtcaa cataggaaga
37441 ccccatctct acaacaaatt taaaaattag ccaggcttgg tggcacatac atagtcctag
37501 ctactcagga ggctgaggca ggaagatcac ttgagccctg gagttggagg ctgcagtgag
37561 ttatgatcac gccactgcat tccagcctgg gcgacagagc aagaacctgt ctctaaaaca
37621 acaacaacaa caacaacaag taatgtgaga aacaaatgcc ctccctccct ccctgagggt
37681 caccccccagc ctgacctgat ggatgggaaa agataagctc agagatgggg taaggataag
37741 gtgggagtta ggaatggggt agaggtcagg gatggactga gaactaggaa agggaaggat
37801 caaagatgga ctggaggtct ggaagcagga agagagaaat ggagtaaggg tcagggctgg
37861 agaccagcag aaccgaggtg aaatgtagga ccatctgatg gatggtcaca ttaataactt
37921 taatttttca ccaccagctc ccactacagc atccaaacat cctgctctga aggagtggga
37981 gcagggacag ggaagaagaa gaaaccttaa ctgagtgaac attttaaaac aagcaggaca
38041 tttcgaactc aaatagtaag agtcactttg ctgtccaaat cccaccagtc cttggtgtaa
38101 tactacaagt gggacttaca agcaggaccc gcaagaaggt tctagtaact acctcagctt
38161 gcaggacgta aaatgcctgg cgaactttgc attccaagga caggagtctt ccaaattaaa
38221 attttaatag tgtgtgaaaa tagttcgcca tttactctta ggatgaggta gtgtttagaa

FIGURE 3-K

```
38281  agtcagcctg gaaatccagg ccaacatttt cagggatgcc tcctattctg ggaggagaaa
38341  aaccagcttg ttgcttccct ccagtagcac aaagcagcag ggaaccggaa agagagaggg
38401  gtctgaacag cttcttctac caggaaacca gccaaaggtc tccgggcatg taaggacatc
38461  tgttttcaga agctatggtt tggattttgg cattcttagg agtctaggag caatcctact
38521  aatttgggtt ctagctccaa ccatactacc tacctgctgt gcgaacccag gcaagtgatt
38581  ttacctctct gagcctctgt ttgctcatct ggccctctag acaagattaa cttggcagtg
38641  ccttgtagct ctaatattca tactctcata agactggatt tgttctcaca ggaatgaatt
38701  agttccctgg agagttggtt gttaaaacac cagggccccc ttgggtttcc ccatcttcac
38761  acgtgtctac tttccctttg accttccctg ccatgttgtg atgcagaagg aaagccctcg
38821  ccagaggaca gggccatgcc cttgaacttc tcagcctgca gaaccataag ctaaataaat
38881  ctcttttctt tataaatgat ccagttcctg gtattccatt acagcaacac agacagacac
38941  tcccacatca tcgtcacaca gggcttgggc tgcctggccc tggccactgc ctgccaacac
39001  caggcctggt ccctccatca tctaggacct tggggaacat ggcaccggtg aagggcatgg
39061  gggttaccag aaagagtgag aagtcggcag gtcagcacgg ccacacaagc caccctctcc
39121  caacttatgt gcaggcatga tggctcccca tccaccagca ctcctgaggc cactacacca
39181  gggaggaaag aagtgcatgg ggaggcctgg tcacacctgc ccttcctttc agaaccctgc
39241  atgtggctga tggcaggtcc tcggcgcctg gtggaagaga aagggaaggt ggcactgggc
39301  tccacaggca gtgattctca gccctgcttg tcactgggag ctggttcctg ggtcccaccc
39361  caatctctta tgattccata gatcaggagg agacccaaga acccaacttt tcacttagct
39421  ctccagggaa taccaatgca aacggccact gaccacccct tgagcgccct gtggatgaga
39481  gggtggaact agaggttcta gaagcaagtg cttgagtccc ctcactagct gaatgactct
39541  gggcagttta catctccttt caaaccctca actgcccatc tgagaaatgg ggatataaaa
39601  gcgtctgctt cttggggtta tttagagaaa tacatgagaa aaaggatgga aagcacttag
39661  cactgtgcct gtacatgttg agaatctata agagaagatc ccctcttact attcaggggt
39721  gacttgtaca aatgccttct tccatagcaa tcccttacta aactggagca tttcataggg
39781  aagatgcgaa taagggcaat gaagcctcag atacagaggg cagcggagtg ctgagacacc
39841  tccgtgccct cccacaacat cccagtgctg tggtccagca acatacccgc cccatccctc
39901  caactcagaa catgctcatt gatgtgcctg aacccaggct aaatgggagc acacctatag
39961  tgggggttcc cctgtcagag atatgcactg caagggcctg cggtccagga cacagaggaa
40021  ggaagagacg ggMggagaat gtgaatcaca ttcatgccag ggtgtgttct aatacgtcag
40081  caaatgtgta ccgaScttcc gggctgcaga ggaaggtggg cgtgtgtggg gttagtgact
40141  tgaagggaaa gaggggcatg aaaaagaagg gggtgaagag ttagaactgg tgcagcagct
40201  gggtctcaca cacaaatggc atgactctac ttcacaggaa ctactcagaa agggaaaaga
40261  ataactctaa gaaacaaaga tctgcctcca aatccaagaa tggcagtaga aaatagggtc
40321  ccccacccag caagcaccca catcacgttt catcagaaga actctctatt ctcagccagg
40381  cctggcacac tgccttgacc ttcccgatcc tggagctggg ccctgggctc agaagacccc
40441  agactgtcat tgctggaaac agaatcattg cagatatagt tagtcaagat gaagtcacac
40501  tggaatagag tgagccctaa tcttacagga ctagtgagtt agtccatttg tattactata
40561  aaggaatgcc taagacttgg taatttgtaa agaaaagagg tatatttcac agttctgcag
40621  gctgtagagc aaccacggtg ccggcatctg cttctggtga ggcttcagga agctacaatc
40681  atggcggaag gtgaaaagga gagcaggcgt gtcacatggt gagagcaaga ccaaggcagc
40741  aagggtggga agtcccagac tttcaaacaa ccagatcttg tgtgaactaa ctgggtgaga
40801  cttcactcat caccaagtgg acggtgctaa gccattcgtg agagatccac cctcatggtc
40861  taataatctc ccaccaggcc tcacctccaa cattgggaat cacatttcag catgagcttt
40921  gcaggggaca aacattcaaa tcatatcaat tagtgtcctt ataaaaaagg gaagtttgcc
40981  ccatgaagga agagattggg gtgatgtgcc tacaaagcaa ggaatgccaa agataagact
41041  gccagccaag ccccagaagc cagcagagag gcctggaaca gattcttcct cacagccctc
41101  agaaggaacc aaccttgcca acaccttgat cttggacttt tagcctccag actatgataa
41161  aataaatttc tgtgaagtta cccagtctgt ggcacttaca gccaccctaa caaactaata
41221  cacccagaac cttactatcc ttcaccacac cgatcaccaa tgaggcaagt cagaaatttc
41281  caaaaagtct ccacatttat ttttctaccc aataaaccag acatgataga ggaattcaaa
41341  tataaacaag acataggccc tgccttccag agctcctagg cacacaaatg ttagtcactt
41401  tacactttgc aaaacacttt cacatatatt gtctcttgta gatcccccca tctcacattg
41461  agatgggaag ggcaagctgt aggttactcc cattctgcag atgagaacac tgagttttgg
41521  agagttggtg gcttgctcca gggcacccag ccctggagcg gcagagaggg agctgcagta
41581  aagtcttgag tccaagacaa cattggtgtg gctgttttct ttcagcctgt cttgctgcct
41641  ctcctgccgt cttggaacct cttgcctcct tcaaacagaa gcccctgcac tgaggtctct
41701  tgtatctaag atctgcccca agtgagctcc aggaggcaag gagggaggcc agccatgtta
41761  aagaccaagt taaagaccaa agtggctgca gcaccatgag ataccagtgt agtcccccac
41821  cttctctcat agtcccttac cctgaggctg ctcccaccag ccccgcttac ctttttccca
41881  tccccacccc cggcccccag cctctcccat cccacaggca gggcaggttc ctggtaacag
41941  cattgcccag tgttgtcaaa atagcgagga gctggaaaga attccagaat gaaatctatg
42001  tgcctctgca gcctcccaac tgactgccac tgtggcagct aaacccgcta atgtgctaag
42061  cgttttctgg cagggccagg aggcagggtt ggtattttgg gagctgggag gggcttggag
```

FIGURE 3-L

```
42121  gcaacatcct cggcaagtct gggaaggagg tggaggcagg gaggcaaagg gaaccgaccc
42181  agaacacttc ccatgacgtc accgagagcc acctgagctg cacctcagga ttcctgaaac
42241  actgctggcg gagcggagag tacggagagt acgctgggcc cggtgggcag ggctcagcag
42301  ggcatggagg ctttgtagct aggcctgggg ggagagggcg ctctagtatg ccacttgtca
42361  tgtacacttt ttctgctgac acagagcagt gtgcaatgcg tccgggaggc tgggcagccc
42421  agggctggcc caagctctgt ttgcctgagt tggcagtggc agggcccatg gtgaccaaac
42481  aaaagacggg aagccaaacg cccagaggct ggcgggtgag gctccctgag agggctctga
42541  ggatacaggc actggcgaag aagggccagg ggctcatagg gataggaatg ggKttggggg
42601  caatgggaac ttgcaggcag ggtcccctct ggtttcccag acacaccgca tcgcctctgg
42661  gcatcagcct ctccacactg tccttgcttt tcgggggctc cttcgaaaat gcttcaaacc
42721  atgacaccag gaaggggagc aaaccaccaa atctgagatc caaacaaggt cccagcattc
42781  ctttgtccct gatgggctYg gctccaaatt ctgggagaaa ccagagggtt ttgagagagt
42841  caggagaggg gttcagggga caaaaaatag ggctggtgaa gaattccttc cgggagccgt
42901  ctggtagcag ggagtaccgt gggtggcaag gccgtgctgc tcaggtgagg ggacatggtg
42961  tcattctggg aacagatgag ccagtaaaac cctgacagga ttcaatctgg aggaaaaaaa
43021  tcccagctga taggaactgc tgagcgcttg cagcttttaa agcccctcta gcaaaaggat
43081  ctgcaaaagY tttgcaaact ctgaccacgc agccagtcag ccacttccca gcccacagca
43141  acgagagagt ttttgctaca tcagaacaaa cccactcaca aatatctcat cctggcctcc
43201  cccccggcac ccattccttt tgggtctgtg tgcctgcctt cctcagaaat gacattttgt
43261  cagagcagtg agagatagtg ttgcaggggc caaaggcagc caggggggctg gactgcaact
43321  aggctcccct gctcactagc ggctgacccc aggcagggca cttaaccttt atgaatctca
43381  ttttttaatc tgcaaaaatg caaataatga agctgctctt acagagttgg tgagaggact
43441  aaatgagatg atggatgctg gggctcgagc acagagatga ggggctccat gacggctggt
43501  ccccctcctt cccaccttcc tctccagggc cctccggcac cataatcagt gtgggaagga
43561  tgcagaacaa tgggagttta attcagaggt tggcgaggca aaggcatact ttaatagaat
43621  tctctctcgc ccagtacatg gtttcttgtc caccccaccc acctttcccc atctctaccg
43681  gctgatagtc tctcagctag tagacctttt cttgtttagc agggccactt tttttaaaaa
43741  ctccagacgg gtaccctcca tgttgcaggc gacgtggccc tggatcactc aactgactgt
43801  catttgaatg gtgcccccag agtgaggcaa tggtgcagcc ctcctaaggc ccttcctgag
43861  tgtccctcct tcatgaagat gattctgagg cttcccaggc cttcaccctt ctttgaaagc
43921  ccatagcagt tcatatgcac ttctcttcta tgctcaccaa actctgcctt catcatactt
43981  gggggatgtg tgtgagagcg agagagagag tccagcaggg attagatctg acttttatat
44041  tctctacagt gcctggcaca atgctaacac atttgatgag gccagaatta ccctgatacc
44101  aaaaccagtt aaagacatta ccaaaaagaa ctttttcaat attttaattc ttacctcata
44161  ccctagataa aaattaactt gagcggggtg tggttgctga cacctgtaat cccagcactt
44221  tgggaggctg aagcagacgg atctcttgag accaggagtt tgagaccagc ctggccaaca
44281  aggtgaaacc ccgtctctac taaaaatata aaaattagct ggtcatgatg gcaggtgcct
44341  gtaagcccag ctacttggga ggctgaggag aactgcttaa ccgggaaggc agatgttgca
44401  gtgagccgag atcgcgccac tgcactccag cctaggcgac agagcgagac tccgtctcca
44461  aaaaaaaaaa aaaattaact cgacatggat cgtagagcta aatggaagac taaaactttc
44521  aaacttcaga aagaaaactt ggaagaaact ttttctaacc ttgggttaag caaaaatttt
44581  cttaaatagg acataaaatg catgagtcgt aaaggaaaaa aattgataca ttggacttca
44641  ccaacatttt aaatgtctac tctttgaaag gcactaatat gaaaatgaaa acacaagcta
44701  cagactgaga gaaaatattt gcaaaacata tatctgataa aggatgttta gtcagaatat
44761  ataaagaact cctaaaactc aataagaaga caaatgactc catttctttt aaggtcaaag
44821  aactgaacag acatttcaca aaagaaaaga tacagatggc caaaaagcac atcaaaaaaa
44881  aaaggttcaa cattattagt tagcagggaa atgcaattaa aatcacaatg agataccgtt
44941  aaacatctat tataatggct aaaatttaaa aggctgacaa tgccaagctc tggcaaggat
45001  atagaataac tacgactctt acacactgcg tgtgggaatg ctaactggta cagccacttt
45061  ggaaagcagt ttggcaattt cttttaaagt taaatataca attatatttt catttctccc
45121  agagatttgc tcaaaagaaa taaaaacata tgtctacaca aagacctata aataggaata
45181  ttcatagcag ctttctgcat aatggccaaa acttgggaac acccaagtgc tcatcagttg
45241  gtaaatgaat aagcaagttg ccatgcagcc acacagtgga acgcacccat acatgcagta
45301  acatggataa gtcttaaacg catggtatca gttcctattg ctactctaac aaattactac
45361  aaactcagta gcttaaaaca acacagattt actgtcttac agttctggag gtcagaacgt
45421  ccgaaatcag tctcactggg ccagcatcaa ggtcagtttc cttatctttt tcagctccta
45481  gaagccacct gcattccttg acccatggcc ccttcctcca aagccagttg tgttgcatct
45541  tctatctaac tgcttaaggg aacaggactt tatcaaggca cagagccaca gcaagtatgg
45601  gaaacatata agtagttcca tatttccagg gtgtgtaatg taacatttcg aagataagga
45661  aactgagatt cagagagatc caatatctta cccaaggtcc aggacacagt gaaaatcatg
45721  ttcaacccta aggcctctga atccaaatcc agagtccacg tgagccctca gacttccctg
45781  ccttgaccgt ggatagtcat ctccaccttg tctcagagcc ctgcactccc acccttgtac
45841  actgcaccag cagctaagcc tgtggagcca ggcagaaccc tgggcctccc cacctcccca
45901  ccctgggacc aacagattcc cagctccagc aggataacat cataacctta ataagattga
```

FIGURE 3-M

45961 catttgcatc ctgttttata ttcatcagca ttcttacact tacgttatgt tatctggcct
46021 tctccatcgt atgagtctaa tggtgcagat attctccacc ccacattctc caaccctcag
46081 aaagctgagg tactttctta tacagcaagt gggtagaagc caagatcaga accttacccc
46141 caacccaccc cttgatacca gaagagccta ttgactcaga acaaaactca ggcaaacccc
46201 ccacctagct gagcttcccc tccttcatct gcaggatgga aaacagggta atacctgccc
46261 tgcctatcct gatacatcag gcataaatgg aaaaaaaaaa aaggaagtag tatgtaaaag
46321 cactttggaa atgttaaaat gctatgcaaa ttataaggaa ctgttctgat tatactctca
46381 agaaatgtca caaattttct ccaaaagtga ataattgtca ttacttttca acttcatccc
46441 cttgaaaagc cagtttctcc tcctgtcatc ttttatttct ggttaaagca gagatgttat
46501 actcaacatt ccaactggaa tatcaagatc aatctttaat tcctccttcc agcttcaccc
46561 tttctgaaat caagaacagg caattctcta ctccacccac catctcatcc tcagtttctc
46621 ttgccatctg ccccctctg aagctagaaa tcaccagcaa ccattttcat tcctcatccc
46681 tcYggtctct cctcagtcag agccaaccta catgatgcaa aagcaccact ccaatcacct
46741 ctctcagagg ctcacagccc cgggggaca gcgtccaggg gtcctccacg ggctgcccca
46801 gtgaagctct ccagcttgtc tcccatgact caccaaccca ggtaagccta aatggccctg
46861 tttcccaaaa ctcaactctc ttcccacttt tatgtcccca ctatatctgg aactcccttt
46921 cccactattt aattagctat gggggtagga agaatcaaac tagaccccca gctcttacat
46981 caacctaaat atattccaag tgaatcaaag aattgtggtt aatggctggg cacggtggct
47041 cacacctgta atcccagcac tttaagaggc caaggtgggt ggatcacctg aggtcagagg
47101 ttcgagacca gcctggctgg gtaacatggc aaaaccccgt ctcaactaaa aaataataat
47161 aataatacaa aaattagcca ggcatggtgg cgcaagccgg tagtcccagc tacttaggag
47221 gccaaggcag gagaattgct tgaacccacc cagcagatgg atgtttcagt gagccaagat
47281 tgcaccattg cactccagcc tgggcaacag agcaagacac cacctcacaa caaaaaccac
47341 acacacacaa agaattgagg ttaaaaaaaa tccagattat atagatgaat atctatctga
47401 tagctgaatg gggctggaat ttctaggcta catagcaatg gagtaaatca gaggagaata
47461 aaatgaataa tagttaacat ttacctattt tcacttcttt taactggata tccatatttt
47521 aaaagtaaac atacccttta cttcatgttg tacacaaaaa ataacaaaat ggatcacaga
47581 atgaaagaaa tgctaaaatt agaaaaattc cagaagaagg ccaagcacgg tggcctgtaa
47641 tcccagcact ttgggaggcc aaggtgggca gatcacttga ggtcacgggt ttgaaaccag
47701 cctggccaac atggtgaaac ccctattaaa aatacaaaaa ttagctgggc atggtggtac
47761 acacatgtaa tcccagctat tcaaatggct gaggcacaag aatcccttga acccaggagt
47821 cagaagttgc agtgagcagt gatcgagcca ctgcactcca gcttgggcaa cagagcgaga
47881 ctctgcctca aaaaaaaaaa caaacttcca gaagaaaaca tggggaaaat cttcataacc
47941 ttggtgtagg caaagattct taggacacag aaaataagaa tcacaaaagg gaaaaataac
48001 ctggacttca aaattaaaaa tttctgctct tcaaaagaca ttgttatgct gggtgtggtg
48061 gctcatacct gtaatcccag cacactggga gaccaagacg ggaggatcat ttgagaccag
48121 gaattccaga ccagcctggg aaacacagaa agaccctatc tctacaaaga agtttattaa
48181 ccaggtgtgg tggctcatgc ctgccgtccc agctacatcg caggctgagg cgagaggact
48241 gcttgagccc acgagttcga ggttacagta agctttgatc atgccattgc actccagtgt
48301 gggcaacaaa gaccttgtct ctaaaaagat aaaaatttaa aaattaaaaa gacattgttt
48361 aagataggag atatggttgt acaacaatgg gagtataatt cacaccacta aatactcaac
48421 atgattccaa tggcatcttt tatgttacgt atattttact acaatttttt taaaagatgt
48481 tgttaagagg atgaatagac aagccacaga ttgggagaaa tgattcatag tacatatatg
48541 tacatagtat agtcatacag tatatagata catgtataca cagtacatat atagcacata
48601 catgtacata catacatata tgacaaagga tctatatcca gaatatataa agaactctag
48661 caatttagta ataaggcaaa caacacaaca aaagatttga acagaccctg cacaaaagaa
48721 gatagccaat aagcacgtga aaatatgctc agcatcatca gtcattaggg aaatgcccac
48781 taaaactaca gtaagctacc gttcacaccc attaaaatgg ctaaaattca aaagactggt
48841 aatgtcacct gctggtgaga atatatagca accaaaagca cagccactgc tgacgggaat
48901 atgcaatggt gcaaccactt tagaaaatag tatggtagct gcttattcta aaaacatgca
48961 tttaccatat gattgagaag tttcactact aagtatttat ccaaaagaaa tgggggaaaa
49021 tggtatgtca aaaaagagac atacacagat gctcgtggca actttattca taatagccaa
49081 aaactggaaa caacccaaat gtccatcaac tggtgaatga ttacaatatt ttggaaaatt
49141 catatgatga aaaactcagc aaaaaaaaag atgttactgc attggtatca ggactgaaaa
49201 aaataaaaat aaaaattagc tgctaatgta cacaacaatg tggataagac ttttttttta
49261 aaaggccgag tgaaataagc tagattcaaa agagtagata gtggtcaggc atggtggctc
49321 acatctataa tcccagcact ttggaaggcc tggggcgagt ggattgcttg agcccaggag
49381 ttcaagacca gcctgggcaa catgggaaaa ccccatctct actaaaaata caaaaattta
49441 gccaggtgtg gtgatatgta cctgtggtcc cagctactag ggagactgag gtgggaggat
49501 cacttgagcc caggaggcaa aggttgcagt gagctgagat tgcaccacta cactccaacc
49561 tgggcgacaa agtgagaact tgtcttatta aaaaacaaaa ggtagatagt gtatgattgt
49621 gtttacatga actcttagaa tcagcaaaat taatctacag ttacagagag ggggtgagta
49681 caaagaggca ctaaggaact tttggggata ataataatat tctattgtgg ttacatgggt
49741 atacacattt gttaaaattc atagaactat ataYtttgaa tgagtacatt ttattgaatg

FIGURE 3-N

```
49801  taaattatac catagtttta aaagaacttt ttaaaagtca agatgcaaac tggggaaaat
49861  atttgccaca tgtatgacaa attaatattc ttaatataca aaatgtcatg caaattgata
49921  taaaaagcat taagacctca acaggcaaat ggccaaaaga cacagaaaag ccacaaaaga
49981  ggaaataaga atatctaata aacaggttaa aattaaagca atgaaatact gtatttcaca
50041  aaccaaatga gcaaatttca cctatcaaat gagtttttt gtttttgttt tttgctttga
50101  tgataatttt tgttttgttt ttactttaat gataatacta gtattgtagg acaagatgag
50161  cactcttctg gtaaaaatat aaattgatac atttctagaa cacaatttgg aacactatta
50221  tataagcttt agcaggaaca ttaaagttaa gacccaaaac cttaacaatt ttcagattat
50281  ttgacccact tctagaaacc tgtgctaagg aataaacaaa gatgaatgat caaaattaga
50341  gatttacagg caagagtttt cagccaggca ctgtggttca cacctataat cccagcactt
50401  tgggaggtca agctgggagg atcacttgag gtcaggagtt cgagaccagc ctgggcaaca
50461  tgatgaaact ccatctctac aaaaaatttt aataattagc ccaggcacag tggctcactc
50521  ctgtaatccc agcactttgg gaggctgagg caggcagatg acttgagacc aggagttcaa
50581  gaccagcctg gacaacatga tgaaacccca tctctaccaa aaatacaaaa attagccgag
50641  tgtggtggca catatgtttg taatcccagc tgcttgggag gctgaggcat gaaaatcgct
50701  tgaacctgag acgcagaggt tgcagtgagc tgtgatcacg ccactgtact ccagcctggg
50761  caacagagtg aaactctatc tcaaaaaaaa aaaaaaaaaa ttagcgggcg tggtggcatg
50821  cacactactt gggaggctga ggtgggaggc agagcctggg aggcagagtc tgcagtgagc
50881  tgaggtcaca ctactgcact ccagcttggg caacagagta ggactgtgtc tcaaaacaag
50941  aaaaaacaag tctggccatg gcggctcagt cctataatcc cagcacttaa caaggtagag
51001  gcaggaggat agcttaagcc caggagtcca agacctgcct gggcaacata gcaagacccc
51061  attctccacc aacaaaaggg acaaaaaaga atttttatta ctgaccttgt ttctagcaag
51121  aaacaatcta ggataagcca ctaggttgag taaacgctgg catcaatatg atgaactatc
51181  gaatattgtc catggtacaa acttaataaa tatttctcaa atatactcaa cgtatacact
51241  atataagaac cttagaaagg agtgagatgt gtagatgact tacacttatg ggtagagaaa
51301  cacataaaca aattgtttca ttgcagcaaa tattggaaat catctaaatg ttcatcaaga
51361  gaggtgaaat taggccgagt gtacttgctc atgcctgtaa tcccagagct ttgggagacc
51421  aaggcaggag gactgcttga ggccaggagt tcaagactag actgagtaac atggcgagag
51481  ctcatctcca ctaaattttt tttttaatta gccaggcact gtcgtgtgta cctgtagtcc
51541  tagctactca gaaggctaaa gtcagaggat cacttgagcc caggagttca aggctgcagt
51601  aagctatgac acgccactgc actccagcct gggtgacaga gtgagatctt gtattgaaaa
51661  aaagaggtga aattaaataa atcatggtat gtccacacaa tgaaatgcta tgcatctgtt
51721  aaaaaacaggc cagatctaat ctatgtccta ggctgggcgc agtggcttgt aatcccagca
51781  ctttgggagg ccaagRtggg tggatcacct gagatcagga gttcaagatc agcctggcca
51841  acatggtgaa accccatctc tactaaaaaa tacaaaaaaa ttagctgggc gtggtagcac
51901  atgcctgtaa tcccagctac tcgggaggct gaggtgggag gatcacttga acctggaagc
51961  cagtggaggt tgtggtgagc caagattgct ccactgcatt ccagcctggg caacaagagt
52021  gaaactccat ctcaaaaaaa taaaataaaa taatctatgt cctagattgt taaaaaatWt
52081  ttaaaggtgc agaacagggc acttggaatg ctactatttg aagggagaaa aaaataaaag
52141  gacacataaa tatttatgtt tatataacat caccagaaag atacagaaga aaaaacaaaa
52201  caggtaagtc tatactgaaa gactagggtc tgaggatgtg aaaacagggg tgacagagac
52261  ttagatttac tttttgctct caattacttt ctgtagtgtt taaatttttt accatataca
52321  ggtttttctt tttcaaaaaa acaccaccaa aaacaaaaac acagattttt ttttagaagg
52381  aaggggagtg ttatggattg agtttcttag aaggaaggag agtattatga gttgagtttt
52441  ttagaaggaa agagagtgtt atacgttgca ttatatcccc tttatatgtt gaagtcctaa
52501  ccccccctaca ggttgaacat cccaaattca caaatctgaa atgcaaaatc ctccaaaatc
52561  caaaactttt tgagcatcaa catgactcat gaagatctca cccaatttcc cagaagccct
52621  cgctcccatc ggactagaat cttcttttga gaggcctttc cttttctcat ccctattgca
52681  accagtacaa aaagcaaaac agaaagacac taaagcaaaa ccaaagtttt agtcacaaaa
52741  accaaatggg ttggcgaggg aaggtggtgg ttacagggaa ggttttaagc tcaccaggcc
52801  ctgtagctga ggtttcaaaa tggcagctag ctggacgcca tctggtctaa atctgctcac
52861  ccaaacctcc ctccagttct ccaagggatg tccctaatta ctagcacagc cagctgccaa
52921  catggaatct aggtctccct tctttcaaat attagagatg tgagaccaga aagattactt
52981  cccaatacca gagaaaaggc tacgagggag gaacaaggga gagaaggaga ccttcactgc
53041  agcatgcaag gacatggcgg acgcgtccca tcccatatct cagctattat gcattcccaa
53101  ctggcttggc caacatccag agagactcca ggctggtgag tctgtgaatg tgaagtagag
53161  ataagcaagg cggcgagggg gagctcgccc actgttatat acgggctgca gcattgcgtg
53221  gcgaggagtc aggggaatct tccatttcca agacatttct gcactcgatc ctgctcccca
53281  agactaagaa ttctttcttg gcttcctccc actccctgga atgtaactaa gcagttacct
53341  tgagaacaag gaaaaattgg gggctaggga ggtgtatgag tgttttaatg tgttaaaaat
53401  cacatttacc ctggatactt gtttgcaata aaacagagag taggaaactc cattcttacc
53461  ttcatctatg ggacaggcac ccagtggcaa tcattgtgag aaatagggag aggctgatct
53521  atgcttaggg agctcagaaa atgctatgct tccttccagc tgtagggaag ctgagccact
53581  ggggaagagt ggatgcattt ggacaaacag gaatggatgg aaggctttct aggagagggg
```

FIGURE 3-O

```
53641  accacatgag caaagcctga agaaggggtg gttcttatct agcaatgagg aagtgtgagg
53701  agaggcaaag aaatatcttc tagaagctga gggatcaggc actggaagct tatctacctt
53761  ctccaactcc agggtaccac caaggagagc aaaatccaga aacaaggggc acctgcaaag
53821  tggtggagga aagaggagga aaggaatgtc tgccccRagc ccagccccag ggtcagatgc
53881  cgggaaaaat tcctgggatt atttggtgct gtatgagatg gatctggagc caccctggct
53941  ccaagaggga tgtgggtttg gactggggct Kccaggactg gtctggggca agctaaggac
54001  actttcacgt gtccacactt ttgtacttgc tgttcaccag tatctccttt ccccccttct
54061  cctccctgtt catggaaatc cttaaaaatc atctcctctg ggaagttctc ctgatgctct
54121  cattcagaat taacccacca ttctttatgt cgcttcaaca agttgcttgt atctcttttt
54181  ccatttacgt tttcttttte ttttttttt tttgtttgtt tgtttgtttg ttttgagaca
54241  gagtctcact ctatcgccca ggctggagtg taatagcgcc atcttggctc actgcaacct
54301  ctgcctcccg ggttcagttc aagccctgct cctcctcagt ctcccaagtg gctgggatta
54361  caggcacctg ccatcacccc tggctaattt ttgtattttt agtagagatg gggtttcacc
54421  atgttggcca ggctggtctc gaactcccga gctcaggtga tccgcctact tcagcctccc
54481  aaagtgctgg gattacaggc gtgagctacc acgaccagcc tacgctttcc tatatgcctt
54541  ttttttatt ttttgagatg gagtctcact ctgttgccca ggttggagtg cagtggctca
54601  atcttggctc actgaaacct ccacctcccg gattcaagcg attctcttgc ctcaacctcc
54661  caagtagctg agactacagg cgcatgccac catgcccaac taatttttgt atttttagta
54721  aagatggggt ttcaccatgt cggccaggct ggtcttgaag tcctgacctt aggtgatcca
54781  cctgtctcag cctcccaaag ttctgggatt acaggcatgg gccaccgtgc ccagcctatg
54841  ccttgtctta taattagtct ccctctcttc taaatgtaaa ctctcagatg tcaagaacta
54901  gatcttatac atctttctga ccccatcccc cagcaccagc tcttccccag agcagacact
54961  cagtaaatac ttattgaatt gaatatacag acccacaagg agcaacctgg tactggcgga
55021  catttttgat tcctgcatcc attcattcat ttgctcaaca accatgtgag tgcctaatac
55081  ctgccaacct ggctcacact ctctcacaga gcacactcaa attgtttcat caaccagcta
55141  aaagtaagtt cagtatcaat tcagggcttc ttttcagttc tagaacatag aaaatttcac
55201  aggatcacct gaacaccgca gggcgggtgc agcctacatc ccattgcctg tttctgtcca
55261  gatggaaagt cacaccctca agaaggcccc tggggaagac agattgtctg ggatgtggat
55321  gagggccccc acaggctctg cccagcccag cccagcccag cccagcccag cccacccagg
55381  ggaaccctag atacacctat cacaggagat ggcagttaag ataactctat ctacatacac
55441  agtaccctgt ggatactaag taaggagctg gtcatttgga atgaggcagg agagctctag
55501  gaagtcctct ctgcagagta gatgtggcca tggaggtggg aggattaact gaaggtcata
55561  agaaaagtct gcccttctca gagatcaata agtgcaaatt acttgagagt ccaggaactg
55621  ctctgccatg gaggaaaaat acagatgggg gcaggggcgg gaaaggaagg gggcaggtta
55681  gggacagtca ttgatctcgg cgggctcttc caagtctaaa attgtaggat tctaacctga
55741  acctcccctg ccaacaactt gaaagagttg tcttgagatc agggacatgt ggagagaaag
55801  ttgaatagca tgtcagataa taaaccaaga agaaattgtg aagtcaaaac aagggataaa
55861  ttttccttat ctctctaaga catgcatcaa agcagctgRc tgctgtctgc ctggcctggg
55921  aagatgataa tggtttttaa tggcccgtaa acaataggac cagcctttgt aataaagttt
55981  aaccaattaa tcctctaggc ttcctccctg cactctgcct actctccgct cctcctttct
56041  gtgatggggc aaccttcagg ggtctgaaaa gcaagggatc cagcccctag acgcaaacct
56101  ataggtgaca ataagagttg cccagaggaa gtgctaaaga aaattgcctc tatcctgatc
56161  tggctaaaag gaggaagaag ccctcttttc agcgaaaagg agaggagaga tgatgcatca
56221  aggaatttgc acaagaggta aacttcattc ttattcagga gctcattcat tcattcagca
56281  aacattcatg aacacctact atgtgcaagg catctgcaga gggtgcaaaa atatacaaag
56341  aaaatacctg cccaagatgg aattaaagcc tggagggggt ggtactggga acaagacaca
56401  ttcacagatg gccgggcgcg gtggctcacg cctgtagttc cagcactttg ggaggcagac
56461  gtgggtggat cacgaagtca tgagatcaag accatcctgg caaagatggt gaaaccccgt
56521  ctctattaaa aatacaaaaa attagctggg catggtggcg cgtgcctgta gtcccagcta
56581  ctcgggaggc tgaggcagga gaatcgcttg aaactggaag gtggaggttg cagtgagcca
56641  agattgtgcc actgcactcc agcctgggca acaagagcga aactccatct caaaaaaaaa
56701  aaaaaaagac acattcacag atatcaatcg taaaagttag tacaatagcc tctgcaaacc
56761  actgtataag tccagcttct ttttccattt cttctttcat gtgccatgtt ccgcttgggg
56821  tggttacgac agacaatatt cagaaaatgg aagaatcaga aagagaagaa agtggaacac
56881  acaagagaaa gaaaaccaag gccctcctcc ctcctgccat ggccctctct gtcttcagcc
56941  ctagttctgg aaaagaaaag ggtgtgttgt cactagaggc ctccaggcag gtccatgcca
57001  ctgagtcaca tttcaagctg tgtttgcttc ctacaccaac aaggcatatc ttttctgaat
57061  tgtgcaatat tctggccacc tgggtttttc tacttggtat accaactcct agtactcaac
57121  aaccattccc tttactcttt ttttttttt ttttttttt tttttgagac agagtctcac
57181  tctgcacaat ctcggctcac tgcaacctcc gcctcccaag ttcaagcgat tctccctgtc
57241  tcagcctccc aagtagctgg gactacaggt gcccaccgcc atgcccagct aatttctttt
57301  ttacttttag tagagacagg gcttctccat gttggccagg ctggtctcaa actcctgacc
57361  tcaggtgatc tgcctgcctc agcctcccaa agtgctggga ttacaggtgt gagccactgc
57421  gcctggcccc cctttactct taaatacagt gtgtgccagc tgtagtagtt tatgcctgta
```

FIGURE 3-P

```
57481  ctccgagcat ctcaggaggc tgaggcagga ggatcacttg agcccaagag ctttaggctg
57541  tagtaagcca aggtcacacc actgcattct agcctgggca acagagcaag accctgactc
57601  ttaaaaaaaa atacaatgtg taatggtgat ataataattc aaaggttaat catctctact
57661  tttgaaaaca ataatgaata ccttgctttg catttgttta gcctcacatt tttttcagag
57721  ttcttcaaaa aatattatct catgttcatg ctacaagcaa ccttggatgg taacaaaggt
57781  taaacatcat tatttccact ttaaattttc cctcctttct ccatttgttt tctactcaac
57841  aaacagtggt tgagtacctg ctatgtgcta agcaccatgc tgggtacagg gtcatgcaac
57901  atagcagaaa aacagcagat cctgccttca acaggttcat agagctatgg ggaggagaag
57961  tgaggacgtg gtccctgtct ccagggtctg gggcatcttc tagaggagaa gattgagaca
58021  agcattaaaa gattatcatg ggccaggtgt ggtgctcaca cctataatct caacactttg
58081  ggaggccaag gcaagaggat cgcttgagcc tgggaggtgg aggctgcagt gagccatgat
58141  ggtgccactg cactgcagcc tgggcaacaa agtgagacca tgtcttaaaa aaaaaaaaaa
58201  acttaatatg atttttaccag gtagaaaaga gaggaaaggg gattccagag agtgggccag
58261  caatctatgc acagagaaga gtgagccatc tggggcatcc agaccatggg gtgaggagca
58321  atgaagggca aaggctgaga aagactgaag agggaggctg ggcccaaaca acggagccaa
58381  gagtgccttg ccaaggagtt tggactttat cctgggggca atgaggaaat cactgaaggt
58441  tttaagcaga ggtgggtgac gtgatcagat ttgtgtttca gcaagattac cctggccacg
58501  ggatggagaa agcactggac tgggaagaga ctgggcagac tctagaaaca ttcccaaaat
58561  agaattcgca ggactgggtg tcccactgga gataggagga aaaggagata atagaaacag
58621  ccttctcact tgggcttgag gggccctcgt ccaatatagg agaccagtaa gaaagccttg
58681  cccaaggtca cctcaagtta cttgatggct aactcgaatc aagaagctca agttccttct
58741  ggtagaaacc tcttagtctg cctcattcac attgatcagc aaataaattt aaccatcatc
58801  tgccaggctt tccttgtttt ccactttata catgtgcact ttttgtccta tggagattat
58861  taagatcatt agggcaaaat tccttttgac ctcccataag gcctctcaca gtacaagaca
58921  atatgtggtc aataaatatt tgttattaaa tcaaactcca ggagccccaa gaacccatca
58981  ccaggcattg ccttacttac aacatgagat atttgccttc tttaatccct aataaacagt
59041  tttttgtttt tgagacagag tcttgctctg ttgcccaggc tggagtgcag tggcacggtc
59101  tcggctcact gcaacctctg cctcccaggt tcaagtgatt ctcctgcctc agcctcccaa
59161  gtagctggga ttacaggtgc ccaccaccac gcccagctaa ttttttgtat ttttagtaga
59221  gacagggttt caccatgttg gccaagctgg tctcaaactc ctgaccttgt gatctgcctg
59281  cctcagcctc ccaaagtgct gggattacag acgtgagcca ctgcatctgg ccaacagtat
59341  tttttattta attcctagat aaagtaattc ctagatgctt gaatttcaca atccaagata
59401  atttttttt gagacagagt ctcgctctgt cacccaggct ggagtgcagt ggcgtgatct
59461  ctgctcattg caacctcacc tcccaggttc aagtgattct cctgcctcag cctcccaagt
59521  agctgggatt acaggcgcct gctaccacac ccagctaact tttttgtatt ttttagtaga
59581  gatggggttt caccatgttg accaggctgg tttcaaactc ctgatctcaa gtgatccgcc
59641  cacctcggct tcccaaagtg ttaggattac aggcgtgagc caccatgccc agccacagtc
59701  caagatattt ctaacaacaa caacaaaatt ggagcgttca atgtggtctt gccaactttt
59761  gttttaacac cattgtttct taaaagaaag aacattgaat acatatcccc aaagtcagga
59821  taatataatt ttaagatgaa ttttccaaat taaatcagag ttcctcaacc tcagaccaat
59881  taacactgga ctgaagaatt ctttgctatg gcagtggagg ggaaggagaa gggggagctg
59941  tgggcaaaat tgcaccagtt cagaaccact gaatttttta aagtacgctt tcttaaaaat
60001  ccagttacac tgtccttata ttttggatca gtgaaaacta tcacagactg ccaccagtcc
60061  ataaacagct tttgggaaca agagactcag aggtgaagaa tgtggaattt gaaatcagac
60121  ttgaactcaa atcccagttc cacctctcct cattccctga ccttgagcaa ggcactgttc
60181  caagcctccc tttcctcccc tgcacagcga ggacaatacc accccatccc agttgtgggg
60241  aggattcaat ggaatagatg tgacgcactt agcatgtgac catcgggaac ggaagtatga
60301  cattattaac attatttcta ccacttctat aacatatgtg gtatctaata caaattaatg
60361  actgctaccc aaatttcaga ctttctgggt caaaaggacc ttacacataa tatagtggac
60421  tttcaataaa cacttaccaa atggacaaat gaacccctg tcaccccgat ctcactagtt
60481  ccttccctga aacccgacac atctgagtcc ttttctcctt tactaaccct ttctccaatc
60541  ctgctcatgg gaattaaagc tgtaaaataa gcctggcgca cctcgggcct ctgccctggg
60601  ctctgtgggt gggagcactg tggaagccgt atcaatcgcc cccacctatg agagcctttc
60661  ttcagggcca gccatgaacg tYccccatgt catcagcatc ttcaggctac tgctgtcctt
60721  cttggatatt taacctggag gcgggccagg gacagaaaaa ggaggtggca agatccttga
60781  acaaaaggag ctataaaagg gcgttggggg aagcaaggca aacggcagat taaacaagca
60841  ggcacctcaa ggaaacgtga cgcgggaggg gattccgagc ctctctgtct gcttacttga
60901  cagcaagcag gtcagagttc ctgccatttc cttctagtag aaaccccacc aagaccacct
60961  ccccgccttg gcaggcaaag cactggtagc tctcccaagg gaaaaaaaaa acaaaaaaca
61021  cacactcctt gttacaaacc agacacagtc ctggacagag cccagctaac aataaatttc
61081  tggggttcat ctcacaagta cattaagtat cttctgacaa tctcccaaag ggctgggttt
61141  ttctggccat cacagggctg tagaatattg agtacaagga gagggcttca agaacagggg
61201  atccaagccc tcgtttcatt gatggggaaa ctgaggctca gagaggggga aggaacttgt
61261  tcaggatcat ccaagtatgc tcccgtcagt Rgcagcactg agccaagaac tcaaggcttc
```

FIGURE 3-Q

61321 agttttggca ctgggttttg ccagaggctt gaaaacatac caagaagtgg aaggtgtact
61381 atgagctaaa tctcactttt cctgaaaata tagatgtgtg ctcacacagg cacacatgtg
61441 cataaaaggt agggaaagag gaaaacaaat gtagcaaaat gctagtaaat ggtgaatcta
61501 agcaaagggg atattggtgt tcattgctct attttcacaa cttttcagaa ggcttgccat
61561 ttttccaaga aaaatctgag gaagggctag gtgtggtggc tcatgcctgt aatcccagca
61621 ctttgggaag ccaaggtggg aggatcgctt gaggtcaaga atttgagacc agcctgggca
61681 acataacaag accccatctc aacaaaaaaa ttaaaaattg ccaggcatgg tggctcacac
61741 ctgtagtccc agctactcag taggctgaag caggaggatc gcttgagccc aggagttcga
61801 ggctgcaatg agctatgatt gcactactgc actctagcct gggcaacaga gtaagaccat
61861 gtttctttaa aaaaaattaa aattaaaaaa taaaaaattt tgagggagaa agataaagca
61921 ctaactagca atctaagata aaaaattttt ttaaaaaaat cacttggaaa agagtaggaa
61981 ttcctccttc aatttcattc agtttcccta gctagggact aatatcctta gaaaaagtta
62041 taagaaatcc agaggttaaa ggggtgggac tccaaacaag taagagatcg ccagtattca
62101 tagcagtccg cagctacagc aatcttagaa tccaaatacg atttcagata ttttgtccca
62161 tcagctccta aaatacatat attctggtca caccatgtat cctctgttgt ttttggtgca
62221 aaaagtaaag taaccatgaa ctacagatag ggcaaatgtc agaatttttc caggggactt
62281 tatcaccttt gcaatacagg tagtttataa atgcaataca atatagttta atttctattc
62341 aatttgcaaa ttctttcaca tgagtaaatc atttaatcaa aatccacatc caacccacac
62401 tcaccatgtt ctctatccat attagtattt ggtttgggcc atactgggtc cataatccag
62461 ccctgaaaat ctatcctggg ccacatatat aactcatacc aatagcaggg cactcaaggc
62521 atgaagattt gagggaacac aacctatcca agaaacctca aggaattcct tatggctgaa
62581 gtaaggtctg aggaaaccat gagcagggga atatgacttg gagagaaggt agaggttaaa
62641 tccaggaggg ctcatacacc caacaaagga gttggaagtt tatccttcag aaaacaagga
62701 gtcactgaag ggtttttaagt aggcagtgat ataattggat cattctggca gagtacagat
62761 ggatgaaagt aactcaaagc taggaagact atcggaatga caaagacctg aactaaggca
62821 gtctgagcgg gagtgaactg ataacacatt aggaagagct attgaccaca gcaaatggac
62881 aaaactgggt gggaagaggg cagatgtgtc tgagagagac agaaaggatc ccagatgacg
62941 tctagcttgg ccaactaggt agacgatgac aacactcacc ctgacgggga atccgggagg
63001 aggcgaaggc ttggcagagg aacgtggtga gttttgaacg tgttgaggtc ctggttggat
63061 atacaaatat ggagctccag aaagaagccc aggatgaata tgtgcattgg gagtaaccaa
63121 cattaatgta cggggatgga tgccaagggt gcagaggagg tcagccagag aattgggggc
63181 aaaacaagaa ctgagtcaag gacagaaact ctgggggaaa aaaaaaattt aagggcagtc
63241 agggaagaag agaccccaaa aaacgttgag gagtagccaa agaggtaaag acctaggcct
63301 gcttttaagt gacccactaa actcctccat ctccacaggc tggtcacacc atgctggggc
63361 tggggaagga ataatcatgt gttaaaataa actcatttct caatgaaaag acaaacagcc
63421 caatttaaaa atgggcaaat aacaaataga catttctcca cagaaaacat acaaatggcc
63481 aatagacaca tgaaaacata ggaagtatca ttcggtcatt agggaaatgc aaatcaaaac
63541 cacaatgagg ccgggcatgg aagctcacgc ctgtaatccc agcactttgg gaggtcgagg
63601 caggtggatc acttgaggcc aggagttcga aaccagcctg gctaacatgg tgaaactcct
63661 tctctactaa aaatacaaaa attagccagg tgtgctggtg cacgcctgta atcccagctg
63721 ctcaagaggc tgaggcacga gaattgaacc tgggaggaga aggttgccgt gagccgagac
63781 tgcgccattg cactccagcc cgagtgacag agtgagactc tgtctcaaaa caaacaaaca
63841 aataaacaaa aaacccacaa tgaggccagg catggagtct gacgcctgta acgccaacac
63901 tttgggaggc caaggcgggt ggattacttg aggtcaggag tttgagacca gcctggccaa
63961 tatggtgaaa tcccatctct actaaaaata caaaaattag ctgggcatgg tggcaggcat
64021 ctgtaatccc agctattcgg gaggctgagg caagagaatc acttgaaccg gggaggcaga
64081 ggttgcagtg agccgagatc gtgccactgt actccagcct gggtaacaga gtgagactgt
64141 ctcaaaaaaa aaaaaaaaaa aaaaatcaga taccacttca cacctggtgg gatggctaga
64201 gtcacaaaga ggtagaggaa ctggacccca catacatcac tgattggaac acaaaatgcc
64261 acagtcactg tggaaaacag cttggcagtt cctcaaaact tcaaacacag agttattata
64321 tgacccagaa actcctcccc taggtatata cccaaaagaa ttgaaagcat atgtccacgc
64381 caagacttgt acatgaaagt ttgtgttagc attcttcata acagccaaaa agcagagata
64441 aagcaaatgt ccatcaacag atgaataggt aaatcaaatg tggcgtattc atacaaggga
64501 ctattattca gccataaaaa ggaagggagt actgaaacat gcaacaacat ggctgaacgc
64561 tgaaaacatt atgctgagtg acagaagcca gatacaaagg ccacatgtgt tatatgattt
64621 catctctatg gaatatctag aatagggaaa tccaaggaga caaaaagcag attagtggtt
64681 accaggggta gaggggacag gggagtgggg agtgtctgct tcctgagtgc agaatttctt
64741 ttttcttttt ctgtcaccca ggctggagtg cagtggtata gtctaggctc actgcaacca
64801 ccgcctcccg ggttcaagca attctcctgc ctcagcgtcc tgagtagctg ggattacagg
64861 tgcacgccac catgccccgc taatgtttgt atttttagta gagacgggat ttcaccgtgt
64921 tggccaggct ggtctcaaac ccctgacctc gtgatcctcc ctcctcggcc tcccaaagtg
64981 ctggaattac aggcatgagg ccaccatgcc aggccaggat ttctttgtag ggtgatgaaa
65041 atgttccggc atcagatgtg gatgatggcg gatgtactaa ccaccactga attacgcact
65101 ttcaaatggt tagaatggtg aattttatgt tgtgtgaatt ttctctcaac agaaaaattc

FIGURE 3-R

65161 aatcatcctt tacaaagtga cccacagcca caaggagctg gggaagcagt ttctccagct
65221 tgcagcagga gctccagata ttctgagtaa agaagccaaa cagcctgctt cccgggagac
65281 ccttaaatgt cttcacacat ccacagctgg gggaggagac gctgatgaac aggagaagat
65341 ggagcatcgg cctcccctgg ctactcactc tgtgctcggg ccgcctcgcc cacttgtaaa
65401 accccaaatg ccagggtggc agcaaggcat acatgcaact atgacttgac ccgcagcgag
65461 gcagcaacca ggtctcacct gggcatagtt accccacaag gacaggaaag gagaaatggt
65521 ttccaggccc aggacaccca gaagccacca aaaaccaacc tcatggttca tcccccttat
65581 ttgactcttc gacacctaaa acctgtgcgt tgtgttcaag atacacccgc cggagaccac
65641 attcccagac atcccgaggg ccagccgtca ccacacgggt gcctcgggga gatgctggag
65701 acacaagtgg ctttgaaatt agcacaactt acatggcctc caggtctgct gtggaaatcc
65761 aggtccaaag aagaagccag tgaggaacag gaggctgatc attgtgttct cagaaggggt
65821 tggagcctct ggggaggggg cactgaggac agagcacggt gttctgcaga tcatatcaca
65881 tctctacttc cctctccagt gatatgcttc cctcactctc ccattttgct ccctctctcc
65941 tgccaaactc ccagggcaac cctggataat aaagtgaccg aatgtgaatg cttccaggtg
66001 gcacaccagc actccagaga aactgtcaga gcaggggcca aacccctggc ctactccagc
66061 cactggataa gcgcaccctc ctccatgcct ccctgctcct gcctgccctg aagggagaag
66121 gagcgcaatg aaactaggtc ttcccttggg gttctgacca cctcgtgggg agattatgat
66181 tgccctgtgg ttgccatgcc tactgtgatg tcactccttt tatcagtcta tctggatgaa
66241 tttttccccc ttttctctga aggaaacctc ctacccagtg cttaggagcc cattgggagc
66301 tcccgccctt gtcttatctc gatctgggac ctcagaagac tccacattcc taacaggact
66361 agctctcagg atgaagaaaa tctctctcaa gctggtcaat ctactcctag atttaagcag
66421 aaccaaacag aacctatttc aacatcccct ctcagaatgc aaacaggaga cggctgtgtg
66481 gaaagatcca atcaccttta cttaattggt cagaataaaa tcacttttct ctttttaagg
66541 cagagtttcc ttctgtggcc cacactggag tgcactggcg caatctcagc tcactgcaac
66601 ctccacctcc caggttcaag caattctcct gcctcagcct cctgagtagc tggaattaca
66661 ggcgcgcacc accacacctg gctaattttt gtgtttttta gtagagacag ggtttcacca
66721 tgttggccag actggtctca aattcctgac ctcaagtgat ccacccgtct cggccaccca
66781 aagtgctggg attacaggca tgaaccacca cacctggcca gaatcacttt ttcagtaaaa
66841 taagtgcaaa gaagaatgag ggtgggggaa aggtttactg atcccacact ttgtcatgca
66901 cagaggaagg caaagaaaga aagaaaactc tttcctaaat gatcacttgt ctccagcagg
66961 tgtatcttga ccattccagc catggatcta tagggagttt catgccaatg aaaagcattt
67021 ttttggggtc ctactatgca tcaggcactg tggcaggtac tggagactta gtcccagccc
67081 ttaaagaact aataatagga tcatggccag gcgtggtggc tcacgcctgt aatcccagca
67141 ctttgggaag ccgaggcggg tggatcatca gaggtcagga gttcgagacc agcctggcca
67201 acatggtgaa accccatctc tactaaaaat acaaaaatta gctgggcgca gtggcagtca
67261 cctgtaatcc cagctacttg ggaggctgag gcagaagaat cgcttgaacc tgggaggcag
67321 aggttgcagt gagccaagat tgtgccattg cactccagcc tgggagacag agcaagactc
67381 catctcgaaa aaataaaaaa agaaaagaaa gaaaaaaaag atcactgttc acaatcacaa
67441 catttacaca gcctccttca aaagaatttc ttctctagaa agataagctg agagaccaga
67501 cagttcgtct gcggccttca tgaagtgaaa ggaggcagag aagaaaggag ttggccatgg
67561 acaggttttg ccccagatcc ttgagtgaag accccagcct gctgcaagct ttttcctcat
67621 tctgtcctcc cctcctggaa tctgtggcag cttccacagg cggtgaaaga aggcagcaag
67681 cagtggctcc agtgccagct gcagcaacca caccagggga atggggccat ctgacacaga
67741 gagggtgtga gggccaggga gactgtccct gtagccagct aacccctcca gtcaagtcga
67801 ctttcacaaa atgggcctgc tcctggaagt cattgatgcc attgttaaaa cacagcacca
67861 ggccgggcac ggtggctcac gcctgtaatt ccagcactct gggaggccga gacgggcaga
67921 tcacttgagg tcaggagttc gagaccagcc tggccaacat ggtgaaactt tgtctctact
67981 aaaaatacaa aaattagctg ggcgtggtgg tgcacacctg taatctcagc tgttcagaag
68041 gctgaggcag gagaatcact taaacccagg aggcggaggt tgccatgagt cgagatcgag
68101 ccactgcact cctgcctggg caacagagtg agaccctgtc tccaaaaaac agcaccctgg
68161 agaagacagc agtccattca cactggaaac atccagactt attacataaa aatggcaatt
68221 agtggccagg cacagtaact cacgcctgta atcccagcac tttgggaggc cgaggtggcg
68281 gtgtaatccc agctactcgg gaggctgagg caggagaatc gcttgaaccc aggaggtgaa
68341 ggttgcagtg agctgaaatc acaccactgc acttcatcct ggatgacaga acgagactcc
68401 gtctcaaaaa aaaaaaaggc aattactgtt ggcatcagca gcacagttgt tttcagccaa
68461 aacttaagac acagaactag ccttatcttt ccagtatatg gtatgtctct tgctattttt
68521 tgcaaattat ctatcatata tattatcata tatattcttt ttttttttt ttttttttt
68581 tttttttgg agacagaatc tcgctgttgc ccaggctgga gttcagtggc atgatctcgg
68641 ctcactgcaa actctgcctc ccgggctcaa gcaatttttg cacctcagtc ccctgagtag
68701 ctgggattac aggagcacgc caccacaccc agctaatttt tgtatttta gtagagacgg
68761 ggcttcgcca tgttggccag gctggtctca aactcttggc ctcaagtgat ccatctgcct
68821 cagcctccca aagagctggg attacaggtg tgagccaccg catctggcca tatatattct
68881 tttaatggca agcgggaaca aaatcgtatt cccattttct gaatgagaaa accaagccta
68941 aaagagggac tcccaagtaa tatcaacttg caaacctatg aatttttatc ttaaatactg

FIGURE 3-S

```
69001  gctttgaggg acctttggca acagcaaagg cattcggcaa tttgtgtgga agtcacatta
69061  ccagcccacc gaataagcca caactatgag ttcttccgtg attagcgggg ctggggatgt
69121  ggcaggcagg cgtggaagtg agatccacgt gaaagaccta agcactgact ccccgaggga
69181  ctggcgaaat gacgcagggg ccagatagat gtcagatttg gccaaagtgt cctcctttct
69241  cacaagcccg tctgcctttc acctgtcctg aggtaggaca accggatgcc ctactttgct
69301  gtgcagcccc actctcttgg ggacagtagc ccaaacaggc tgctcttgct acacacagat
69361  aagaagcttt gccacccaaa accaagaaag gggccggtga ggacacacag ggaatgctca
69421  ctcagtggtg gctgagtttg tggaagtagc aaaacctgca ctgtaggacc cacacctgtg
69481  agagagggcc ccacgggaat ctcagcttgg atcaaagcag gcgggtgctc actccactga
69541  gggcagcaga gcccagctcc aggtgcgatc aggaagtgac ctgagaagct ggagttctgg
69601  actcacagac ccaggcttga atttcagcac cactcagtca ctgtgtggtc ttgagtgact
69661  tgcccttgtt tttcctcatc cataaaatag ggataaaaat acctgacctt gaaaaattat
69721  acgaatgcca gcacacagta catacgcaca gcagagacag gagctagaat tattgatttg
69781  ccaaatgcag cagtgtctgg actcagccgc ccacttctga ccaatggcct tcacttgtaa
69841  cacagctgat ctggagggta aagcctgcag ccgtggaacc aagctctcgc caaggtaact
69901  gacccatgtg gggctcagag ccttgttctt actagaacca ctctagtttg tttgattttt
69961  tttatttatt tatacacata attctgtgca ttcctttatg catatgtata catatatatg
70021  catatgtata atttattcaa atagcatcat atcactgata ctattctcct acgccctacc
70081  ctcccacatg gctactcaca gcccctgcct ttccaggcgg cccgtgttaa catcttagtt
70141  tgtatccttc cacaattctc cccatattca tgtaatttta cacacacaca cactgaatag
70201  tgaagttttg gacagtgtgg gggcgttatt gtttaacaaa aatggaatta tacatcatac
70261  acagtctctc cctctttttt tattctacca taatttgcca aaatcctttt aatggttctg
70321  ttataattca tgatataaag gtgtcaaagc tattcagcca ttcctctact gatgggaatt
70381  cactcagttt caacttttgg ccactaccaa caaaaaagca ataaattcct tgacagtatg
70441  tcctaaccaa agaccatttt ccaaccagac aaaaggccag gtcccaggat gtgagggtca
70501  cccatgaaaa ggacctcatg atgaactcct gattcagcca catcatcctt aaaattctcc
70561  ttagaagaga gaccctctgc tcctagaaat acaaattccc aggcatgggt ctggcttcca
70621  aactctagct cctccagttc ttactgcatc aagtcaattc tgcaggtctt cattttccca
70681  tctgtaaaat aaaatcataa ctgccttagg cggcaggatt tgagagtgct ttagcaaggg
70741  atttttatga gggagcagtg tgtgaaataa tatttttgct gtaattctgt gagatgttgt
70801  ttctagcttg tttgggggtc agggaaggga gggcaaaaca gaactaacag attcaattaa
70861  gttccctctg agactcaagg tccctattta taaggtacct ggggctggat gggaaggcac
70921  ttctatcttc tgcatttctg aaacacctgg tactgtggtt gtttacaagg aaaacttagg
70981  aactggacag gggagctttg gagtgtctac ttgcccagtt ttgtcagatt ccaaagaggg
71041  gcaactataa ttgaaattat aagaagaatg gtccaatcag caaattttgg aaaatctgta
71101  atctcactgg tgtttgattt tcactcaaaa attccctaag ctctacattt cattcagtaa
71161  aaagttgggg ttaggggaga gatggagaac ttcttttgca ggggtatagt gatttatttg
71221  gtctattctt gaaaaaaaat gagaaatttt gggaaaacaa tcttaaaatt tccatttcga
71281  gtagcagttc taaaagactt tcaagagcct ctctgaaaag aagcagttac agcaacttga
71341  atgaatttac ctggttatca tatggctcac atgcatgaga gtgggggaga agacacggta
71401  aaatcaacag agttcgtgtc atgatattct gctttaaaac ccaaatcaag gccaggcaca
71461  gtggcttatg cctataatcc cagcactttg ggaggccaag ataggatgac tgcttgaggc
71521  tgggagtttg agaccagcct ggaaaatatc atgagactcc tctctctaca aaagaaattt
71581  ttttgtaatt agctgggtgt ggtggtacac ctgtaatcct agctattcag gaggctgagg
71641  taggaggatc gcttgagccc aggaggtcaa gcctgcaagg agctatgatt ataccactgc
71701  actccagcct ggatgacaga gtgagaccct gtctctaaga aaaaaaaaaa aaaaacctca
71761  aaacctgact atagaagaat tttcttcatc aaaatcgaat caagcagata attggcaccc
71821  ctttggaaaa ggattcagca gcctctactg aacatacgcc tgccccatga cccagtaact
71881  cccctgcaag gtgtatatac ttaaggtctc caaaagacaa gtaccagcat gttcccagca
71941  ccactactca tggaagaccc agaccaaaaa ctaccacatg ctcatcaaaa gtaaaatggg
72001  aaaataaatc atggtgaatt tacataacag agcactatat agcgatgaga atgaatgatc
72061  tacaaccaca gccaacgctg ggtgaatctc acaaacgcaa tgttaaaggg gaaatccagc
72121  tagtaaagtg aacgtgctag aagattccgt tcgatggggt tagactcaga atagtcaccc
72181  ttggtgggat ggggtaggga gcgttggggt gcctggaagg ggcacggcta ggcttctgtg
72241  atgtttcttc atctgggggt ttgtgacaca agcatgctca acttgtgaat tttgggtggt
72301  gaacatatgt gcacttttg gtatgtatgt tatacttcag tttttaaaac ttaagaaatt
72361  agccaatgaa caaaaataaa catgattgtg agctgggcat ggtggctcac actggtaatc
72421  ccagcacttt gggaggccga ggcaggcgga tcacgaggtc aggagtttaa gaccagcctg
72481  gccaacatgg cgaaaccctg tctctactaa aaatataaaa aattagccag gcgtggtggc
72541  agacgcttgt tatcccagct actcaggagg ctgaggcaag agaatcactt gaacccagga
72601  ggcagaggtt gcagtgagcc gagatcatac cactgcactc caacccgggt gacagtgtga
72661  gactccatct caaaaaaata aataaataaa taaacatgat taaggataaa gcaatccagY
72721  ggacaaggct cttggagcca tattttattt gKcatgtatc attccgggga gctctcctct
72781  gtcaatgcca agactagcta gtcgagtgtg gagaaaaggc attctgtgag aacagatgaa
```

FIGURE 3-T

72841 agggaacaga gaaactggca ctttctttgg aaaagagttg tttccaaaac ctagatgtgc
72901 agcctctacc ccaggtaatt agtccaacac actatccatt acagctgagt ccgggtttgc
72961 ttcaaaagca tcgggggttt tttgggtttt gcttccctta aaaaaaaact tgttttgaat
73021 gaggggaaaa aggcgtagtt aatttttata cagaaatctg ttaaagagat ctgatcaatg
73081 gtgacaaaat ggaagccaga gatggaaaag attcactgtg aggctctggc actgaaggcc
73141 aacttgggct tccattcagt tcaataataa tatttatcat agctgctaac aatttttttc
73201 tcttttttga gatggagttt cactctggtt gcccaggctg gagtgcaatg gcgtgacctc
73261 ggctcactgc aacctccacc tcccgagttc aagtgattct ccggcctcag cctcctgagt
73321 agctgggatt acaggcatgt gccaccatgc ccagctaatt ttttatttt agtagagatg
73381 gggtttctcc atgttggtca ggctggtctc aaactcccaa cctcaggtga tccacccgcc
73441 tcagcctccc aaagtgctgg gattacaggc gtgagccacc acacccagcc aatagctgct
73501 aagacttacg caggtttaca tatgctaggc attgctctct gagcttatat atattaactc
73561 attccatctc tatctatgag gttgggattg atataatccc cactttacac ccagagaggg
73621 taataacttg cccaaagtca cacagctaat aagtgacaaa gctaggcagt ttggctagag
73681 tctgtgctct taacaattat accatattgt ttcctagtaa gcatttctta tgcccgtact
73741 atgtgctggg caccatgctg gacaaaagga gtactgagat atataaaaca cgggtacaca
73801 cataacaggt tctagtccat gcaatgatga gaatcataca ggcattctga ggtctaattt
73861 cagaacaaaa cgttagaaaa ctgcctccac acatctggac attaagacta aaatgcagtg
73921 cctgtttttct cattggtatt attctattca atagagtcaa ttttaagact aagcattcct
73981 tgacacccca cccatccttg ggtcatgaaa ctcaataaag ctctcttcat tttgattgga
74041 ggggggaaaa ttccccgtaa atatgtttgc ttctttttcc tgaatcaccc tcttgtgtga
74101 aaaggattgc ttaccactga tctaacaggg tcttgaatac agggtccctc agagggtaca
74161 agaagcccag gtgttcaggt tattaaaaaa tgaagaaatg caaagcagga ttacaacatt
74221 tttggcacat tatagatcct cacacaacaa atttacattt tcacacacac tcacgcacaa
74281 acacacatac aaagtcaaat acacaatgta aaataattat gatcacaaga ggatgatagg
74341 ctctacattt ttaaatattg actcagaatg ccctgggaat gacgtggtca aggcacagac
74401 tgtgcatttg acactttgtg atgaacatct tggtgcacat ataatctagt gcacagataa
74461 tgtcaaagtc tcattagaat cctccatgga accaggtatg gtggctcata cctgtaaatc
74521 ccagtagttt gggaggtcaa ggccagacga tcacaaccag gagttcaaga ccagcctggg
74581 caacatagtg agtccctgac tctacaacaa atttaaaaat tagccaggcg tggtggcacc
74641 cgtctgtagt ttcagctact tgaggatcac ttgagcccag gaattggagg ctgcagtgag
74701 ctataatcgt gccactgcac tccagcctgg gtgacagaga aagaccttgt ctctaaaaag
74761 aagcctccag gaatgggaaa cccaagttaa atcagctccc actccaccct aactcccttc
74821 ctaagccttc cttaagggaa aaagaacaaa atactgagcc cagtagtaag gaaagaacag
74881 tgaaagcccc agttaggaaa ccatgaaaga gaagaaatga ggtacacagg gagaaaacgt
74941 ggatgaaaac cctagactat ggccaccaac agacccaagc tagccaggtg tagacctctg
75001 tgacacctag aggtcaagtc caagaggtca gctgaagtta cgaaaggcag gcaggaactc
75061 aatcagctac ctacttgtcc acgagctagc cctccccagg cctggcatct atgcttacct
75121 cctggcccac tcctgctggc ctctccacag gcacaggtga ggccctaaac accaggaatc
75181 ttctaatgtg tgctgccttc tcccacgtct gtcccatttc ctaggagtta atcatggtgc
75241 ccattgcaac atggcaccca atgccacttt ccagagtagt gtgacgtcca attccaagac
75301 ctggaggata aaaggtatta aagctcctgg gaccacagaa gagacagggc agaaagctgg
75361 tggaaaacca tctccaacat ctgtttcttt ttctgttaaa atctatccac tgtgtgaatt
75421 gccaaggaaa gttaagattc ctgttctgaa acgagaaggg ttctcttttc acatgacacc
75481 tgcagaagat gggaaacatg ataccccctg cagatcagag gacaccggct ccctgtaaga
75541 ggatacccca acccttgtca gtgcagcacc cggcagggcg atttttacat ttttctcctc
75601 ttgctgaagg gcctccagaa aggccaccag ggctttcccc aagagtaaaa ttctccccag
75661 ctccctagga ggtctcactg cccaccaccc ctttatttag ctaagtagca gagtttttat
75721 tttactgttt attttatttt attttatttt attttatttt attttatttt attttatttg
75781 agacagtctc cttctgtcac ccaggccgga atacaatggc atgatctcag ctcactgcaa
75841 cctctacctc ctgtgctcga gcaattctca tgtctcagcc tcccgagtag ctgggattac
75901 aggcacgtgc caccacaccc agctaatttt tatattttta gtagagatgg ggtttcacca
75961 tgttggccag gctagtcttt ttttttttt ttttttttt ttgagacgga gtctcactct
76021 gtcgcccagg ctggagtgca gtggtgcgat ctcggctcac tgcaagctcc gcctcctggg
76081 ttcacgccat tctcctgcca cagcctttgg agtagctggg actacaggcg cccgccacca
76141 cgcccggcta attttttttgt atttttagta gagacgaggt ttcaccatgt tagccaggat
76201 ggtctcaatc tcctgacctt gtgatctgcc cacctcggcc tcccaaagtg ctgggattac
76261 aggcgtgagc caccacgccc agcctaggat agtcttgaac tcctggcctc gagtaatctg
76321 cctgcctcag cctcccaaag tgctgggatt atgagccact gcacccagcc agtagcagag
76381 tttttaaata acttccagag ccctttactg gagacacacc caagaagatg ggctcttatt
76441 cttcccacag aggactatag accatcaact aaattcactg gctcctactc aagctataga
76501 aagctgagga ggaccttgga acatcctctc taaattcctc ccctagttcc agtccaggta
76561 aacaaaaatt gattctattt tttcaaaaca ccagcccctt cagatggcct ggttaatctt
76621 cattaacatg ttatcttcat taacatgttt atcttcatta acatatttct cagtgataag

FIGURE 3-U

76681 aagggcccag ggttcaacat cccactttgc atccatgcag acaagggctg tgggaagctg
76741 taaaactgaa gaggcggtga atccaggagg gattccccag atatccgtga gttggtgttt
76801 gactttcttg aacacgtctt tcctttactt tacacttaca cttactgagc attcactctg
76861 gaccagaccg cgtcctaaga gcttcacgtg tgtggtcatt caatcctcaa accaatccta
76921 tggggtaggt actggagtta agtgatggcg ctgggatttt aacccagaag tctgattgca
76981 aagcccaact taccatgagg gccacgacat atgtctctac ttacggattc caccagtact
77041 cttgacagga gcatttacct cactatatta taagccttga tctacttttc tatctccctg
77101 aagcataatc agccataatc tgctaaccgt gcctcagttt tcattctcct ttctattttt
77161 ttttcagcgg gggcagggga gacagaatct tgctctgtca cccaggctgg agtgcagtgg
77221 ctcaatctgg gctcactaca acctccgcct cccaggttta agtgattctc ccacctcagc
77281 ctcccgagta gctgggatta caggcacgta ccaccatgcc cagctaattt ttgtattttt
77341 agtagagacg ggctttcacc atgttggcca agctagtctt aaactcctga cctcaggtga
77401 tccacccgcc tcagcctccc aaagtgctgg gattaaaggt gtgagccacc gcccagcctc
77461 attctccttt ctttattttt atgatagcct ctaccttgag tttcacccaa gcctatggcc
77521 atctgctcca aactacattt cccaacatcc cttgtggcca ggtgtggcca tgggcctaag
77581 ctaagccaat gggatatgag tggaagctcc ttaatgacaa gctacttgcc tttggatttc
77641 acctcctgct tctcacatcc tgaaatacag acttaaggat acaaaataga ctaaaacctg
77701 ggtccatgaa tgatgacttt ctggacctac cttacctgcc tggtgtgctc gccaaacgac
77761 ttatttgaca aataaataaa cttctgtctt atttagcttt tttaattttt tttatttttt
77821 tttggacaca gagtctcgct ctgtcgccca ggctggagtg caagtgcaat ggcgcgatct
77881 cggctcactg caacctctgc ctcctgggtt caagcgattc tcctgcctca gcttcccaag
77941 tagctgggat tacaagcatg caccaccatg cccggctaat ttttgtattt ttaagtagag
78001 acagggtttc actatgttgc ccaggctggc ctcaaattcc tggcttcatg tgatccgccc
78061 gccttggcct cccaaagtgc tgagattata ggcatgagcc actgtaccta gacttatgta
78121 gccattctat ttggggacct ctttgttata gcagcttagc ctgtatgctg gattcagcac
78181 ctgaagtccc acagcttgtc ttattctatt caacatctcc agaacccttt cttagggctt
78241 ggaacatagt agttgctcag taaatatttg ttagctgaat gaaaacacca ttatgtaaag
78301 caggtagtac cagggctaaa caaactgatt tactcaagct gtgcaggtgt cccgtatcca
78361 ggtgactaca atgtttcctg ctttcctgga aaacaggagt tcaggacaga gagagtaggt
78421 ttaggaaagc atttgcaata ggaaaggggt aagctttcac tgcatctgtt gttggctctc
78481 tgctgagaca acagattttg ggcttaatca aaagttagct gaggttactt cttgtctgat
78541 aaaaataatg ggctccattt gagcacttac tgtgtgccag gttctgtact aagcgcttca
78601 catacattat ggcattcatt ttcattcaat tctcacaata accctacaag gtagaaaata
78661 ttaatatctc cagatgaaca aaccaggccc ccaaaaaagt taaataactt gccctagctg
78721 catggctcgt tagtgacaaa agtaaaattc aaacccaagt tgtttgagcc taaacttact
78781 ctcttcataa ccaatctgcc actctgcttc tcttcctgtg tgacaagcca atattcccca
78841 ctatggttat aagtttcaag ggctggacca cgttttgtgc ttatctcaat agacgccaca
78901 ggcacggagc ttccactggt ggctgtgaca cttcttctct gtgcatgatt ggtgaggggt
78961 tgcacacacc caccaggaga catgacggca gagggcaaac actctgagac ccagggtcca
79021 cccagctctc ttttgccctc tcctctccct gggctgcccc ggctgcctcc ttgacctctg
79081 cactcttctg cccacgctgc tgccgcccac atcttggaca gcatcccccg gggccgcctt
79141 cctgcctgtc tttgtgtaca attcctgcca ggcccggcag ttctcctagc taaagccggc
79201 aaggtccaga cctgcccgtc tggcctggta atggtgtggt ccacatttca cctcccacct
79261 tcaccctaac ttcatcctca tggccattct ttatccacaa caactccttc ctcccattc
79321 cctagactca gcaaagcagt caggagctag gcaaggtggc ctttatccca gaggttggga
79381 taaacctagg gctcataaat ctagggctgt gagccaaata cggcccctgg gattgctttc
79441 tggttgtgca ctatgctttt tgaaaactga gttagctagg tttaggctgg gcacagtggc
79501 tcatgcctgt aatctcagca atttgggagg ccaacctggg tggatcacct gaggtcagga
79561 gttcaagacc agcctgacca acatggtgaa accacatctt tacttaaaaa atacaaaatt
79621 agccaggcat ggtgtcatat gcctgtaatc ccagctcctt gggaggctga ggctgaagaa
79681 tcgcttgaac ccaggaggtg gaggttgcag tgagccgaga tcataccact gcactcctgc
79741 ctaggcgaca gagagaaact ccatctaaaa aaaaaattag ctaggtttaa agtcattaga
79801 ggtcatataa aaatgcaaat ttcaaacttc tctcagaaaa tcaggagccg tgacaacact
79861 gggctcacct ccccatgtgg caaatggctt gggttgaatg tgaacatcct ccctggaagg
79921 gcataaactt cccagttgac cacagtacag tcccccaacc cccacctccc accaactccc
79981 tattcttcgt ctcctacagg tttacatggc ctgactagcc cccagaggca ttggagtttg
80041 acacgcccac agcagtgcct ccattttgcc gatgagcaaa ctgagaaaag gcaaagagaa
80101 attagtcccc cagatctcca agtgagttca tcagcaaatc cagggctggc aatcattaga
80161 caatatttat ccagctccca gatgtccccg cccctgctct gttcagtcac tgcactcatc
80221 catgtggctg ccccaatgcc ttgggcatgc cagagagaca cctcattctt tcaacacaca
80281 ggtcacaaac atcttttgtg ccatgacaat ctgattctcc tcacgtcacc ccaagattaa
80341 ctcttttcaa atctgtggtt cagattatac accctgaagc tgctgagggg tgacctctca
80401 ataaggctaa tctcccctgg acctaaatac tggtggggcg tgtgtgcatg tgtgcatgtg
80461 tgtgtgtgtg cgcacgcgca tgtgtgtacc catgtgaact cgctcatgca tgtgcttcta

FIGURE 3-V

```
80521  gattaacact ctttcttctc attaacttcc cagacaggtg agaagtcaac atttgcccca
80581  gttagagaac cagaaagtat cagtcccact gccccatcta aaggggcagc agctgctgat
80641  ggttgccata gtggaatgaa gattcagggt tgccagatct ttcaactttt caagaagcca
80701  ggaatataac ttagaagagg ggagcaatgg tctctaaaga cctgtgctga gaaacaagtt
80761  cctaagtgct tgggagaatt tatgaggacc aggtgctaga gaatgccggg aaatgaacgc
80821  tccccttga acgtcccatt ggtgaggtca gaactatgtc ctcaatccct catcagccta
80881  tcccaggact ctgaagagga ccccagggtg cttggctggc tttgccttcc agcctaagaa
80941  tgtctttgac taatattttt cttaatccct gtttgctctg tcctcaaaac aagccttttt
81001  ctctttttct tatccatctt ctccatagag ccctcaggtt cttaactcac acggcactat
81061  ctggatcttt ccagtaagta ggttgccgtt tccatttctt tgattaaaca actctaacaa
81121  agaattcatt gagatgaaaa tgcttatcat gaagcaggta acatcctttt cctccccaaa
81181  ccacaagcaa ttaacatata aatggcctgg caaagggact tctttgagag gtggcagatt
81241  ataactactc ttgctcctac acggataaag ctttttttta aagacaggaa agaacaattt
81301  acaaatgaga aaatgtcaca gatggccaca aactaacttc ccgagggatg agtattttga
81361  tgcagatact tcttcaagtg aaaaatgtac aatgtacaaa atgtccactc gaggatctgt
81421  tcctccccgt gtgggtaatc gccctcctac cacgggcaac gggtaggcaa ccgatatgcc
81481  cagaaaggcg gccaatggga ggacagctcg cccagcctcc cgattggcca ggcccgccac
81541  cagctggaac catggcgagc tttgaggaca gcagagccac ttgatgggat tctctctctc
81601  ttcaaagcat agagcaaaca tcaattaatc ctcaaaagac gctttcaaga agaacattat
81661  ccctttgga gacgcagaag ctaagaggtg ccgcaggtag gtgaagggaa ggcaggaagt
81721  tgaagaggag aggagaaagt ggagcaccgg gctttcagcc ccagttctgg ttctcctgag
81781  agagtcaagt tgcaggagaa gaatgaacct gcagaagccg ggcagggaga ggtggtggag
81841  aagccgctgc tcttgggtgc gcccctcaag gcttcccact gccccatccc ctaggagacc
81901  acggagtgaa aaaagttgga ggcctaagta cagccgtgtg tacccagccg ggcaccagag
81961  ttaagacaac tggacttgtg ctcaaaaacc cttctttagt aaaacaaccc gtaggagtct
82021  gtgggaggag gaaaaattga agtctggatg caatttgcca cactgagtct gttctggtaa
82081  ccgcccccaa cacactcaca ccccttgcc tgaaatgggg agaggggtgg gatcggtgtg
82141  tgtaaacaga ggcactgccc tgcctggctg ctctttccat cttaggagaa ccaatccaga
82201  actttctacg gctcacactc tgatcacagc ctgtcgccat gattggcaat aaacgtcaca
82261  cattctcacg ttatacattt attcgacctt taatattgtg ctctgagatt cctggaaaac
82321  ttgtatgtca ttggtcaaac gttataaaaa cagcccacta ttttttcccc actcacatcg
82381  tccgtgttga ctgccttaca caccgatttc tacatggaat gtttttagcc tgcattttaa
82441  cacttagaat aactttcttt gccacacact acatacatac tgtttgcatc aataaacaga
82501  gcccgtgcca catgcaagtt ataggtttat tggagacatt taagtgttta tagaaaggac
82561  ttttctaaag aatctgaatg ggagttctct caatttacca atagtcaaag aaataaagaa
82621  aattacccta cagctgattt ttttttaatt aaactaccaa caactattct gccatagcct
82681  tgcgtgtgtg tgtgtgtatg tgtgtgtgtg tgtatgcatg cttgctgagg catagaaaat
82741  atctggaagg atttacaaaa actccttagc tttggggaac agaatggaca gacagagaga
82801  aaacctcttt ttttcttgtt ttatgttctt tactgtgcat tgtttacaat cagcatgttg
82861  ttgttttttt ttttaagact agtcaagagc agcagtgaga aaggggaag gaaagaacaa
82921  ggagttcaat ctgtaactgt gaacaatcaa ttgagataac tcactaccct cagactagcc
82981  accatgttgc tttttaaatg tttcaaatca caaaatgaaa aagttactct aagatcatct
83041  atttttccc caggtgtggt attttgtccc catgcctgtg caaagtaagc cagtagaatt
83101  tagaaatagc ttgctgcctt ttttttttt ttttttttt tgagatagaa gcttgctcta
83161  ttaaccaggc tgcagtgcag tggcacgatc ttggctcaac acaacctctg cctcccagat
83221  tcaagtgatt ctcctgcctc agcctcccaa gtagctggta ctacaggcgc ggaccaccat
83281  gcctggctaa tttttagtag agacggggtt tcactatgtt ggccaggatg gtctcaaact
83341  cctgaccttg tgatctgcct gccttaggct cccaaagacc tgggattaca ggcatgagcc
83401  accatgcccg gcctgctacc ttttttacat acacaaaata attgcaaact ttcactgcag
83461  tactgccaac cctcatgtca ccatctacaa tgtaccacgt caagaaagga caatatcttc
83521  aaacccactt accagaagaa aaggttgaag ctgtatttgg caggatactg acatacaaac
83581  ctccaagatc tgggaaacaa ctcaggctcc ttgacttgac ttccttaacg atgacttgag
83641  aaacaccaga aagtccagaa cacataaact catgaaacca accaatagac tggaatctct
83701  cctgttatta aagtcatctt tttggccagg cgcagtgcct cacacctgta atcccagcac
83761  tttgggaggc caagatgggc ggatcacttg gggtcaggag ttcgagacca gcctggccaa
83821  catggtgaaa ccctgtctct actaaaaata caaaaattag ccaggcaagg tagtgcatgc
83881  tatagtccca gctactcttg aggctgaggc aggagaattg cttgaacccg ggaggtggag
83941  gttgcagtga gccgagatca cactctactc tctctactct actccagcct gggtgataga
84001  gtgaaatcct gtgtcaaaaa aaaagaaaaa acaaatcatc ttttattctc ttaaagaatg
84061  aatttttcat gggccatcaa cttcttcagt catttaagta tttgtttaga gaatatttca
84121  gttagaaact caggaaatgg agacctgacc tcaaagggcc ctttccagtc ataaagtaat
84181  tagaaataaa ataaaaacag gttgaagtga taagctgata agaaaatcca atttgtttcc
84241  aaaattgttc acttacagct gtttttctca tgaatatagc tatagttgtc atcagcaaat
84301  cagtggcttg gaattcgaaa ggagtgaaaa tatacatcct gggtctcttg aactcctagt
```

FIGURE 3-W

```
84361  gtttgccaga gctaggcagc aggatcaaat ggtcaatttc agcctggccc tgaaatcaaa
84421  accctctaac agtcacagaa gttctgcagc tagtcagtta cttactgccc acacttcact
84481  cagcacaatc cttaaagtca ccagagggag cccaaatttg cgacaaacca gaatgctttc
84541  cttttagggc atagttgtcc ttctgattct acttttgccc tctcctttcc aagactaatt
84601  tagaaacgaa gggaagtagg ctaagaaaat cctcaccttt aatattaata agacttgtgc
84661  agtccccgga tgatcagagc ccagactaca taggacttga aagaaagaat gagactttga
84721  tatggaatat gaaaaagtaa caggtaggag gggtggagag accaaactga gaggcctagg
84781  gcccagcttc tgcagctctg tcactgactg gcccggtgac gatcacgcct atttactaag
84841  cagggatgac ctacgcttac ctgtctacat ttttatctgt gagatgaacg aaatgaaacc
84901  tccaatggtt tatttgccat gacaattttt aaaaatcagt aggaatgagg gaattaggaa
84961  ttgagcccat gctatacctt cttttcaaaa gagatttcac atgtaccatc tcatctgtat
85021  tcctggagag atagtagccc attatgagtg gctccaattt acagaaaaag taactaaggc
85081  acagagtggt gatctaccca ggacagcatg gtctgtagaa gaagcattgg tctgtttagt
85141  gcattgaaat gtattccaga gatattggtc tgtaaagacc ttcccagaat gtgtctagaa
85201  ggttcaacca cctggtggaa gacatagaac gccttgtcat gccttcccca ttccagcctg
85261  gagttctctc catggtctca gcctgcttct atgaagctta gttttatagg gagcctttta
85321  accaagccaa cctcagagtc ccactggcca aggggatggt gggggaggca aaggattcta
85381  ctggggcagt atgtaaccca catataagtt ccatccaaag caacgcttcc aaacctagct
85441  caccaggaca ggagaaaggt taccatcagg gtactcttag ctgagaactc atacctttcg
85501  ggctttYatt ggcttgaggc atgagtgatc agcagagccc cacatcacaa tccctgtact
85561  cagatataaa ttatattagt gggcagggaa tgtcaggaaa cttaggttgt agaatctaga
85621  tcaacaactg cttaccgagg agctgggtga attgtaaaga aagaaagaaa caagtgttga
85681  atgactgtgg cctcattaaa catttctgtg gcctggcctg tgctttgggc tttcccagag
85741  ctgtttcatg caatagcttt gggcagggca cttctttgga cttactccaa ctacacaagg
85801  aggaggagga tgtctatggt tctttccacc ccagacacag tatcagtcac agccccacct
85861  taattccact actccaattc tgttctgagg gtatcacagg atgctgtggt gggttcccca
85921  gccccacaaa atccttttct cctctaaaac aggagaacat ctttccttcc tttctgatca
85981  tccatcttgc agccagcagg actcccagcc agaacattgt caatttactc agggatccat
86041  gaggcacctt ttggaagatg agaggggttg ggtgaaaaaa ggaagaggaa gtggggtgct
86101  ggtacttctg gaaacaagcc atttctcagc ctgctctgtg aaggtccccc aaaccctaca
86161  cagatctagg aagggaaact gcattccaag aatgcatggg ggctagaagg acccgctgga
86221  gcatggatcc cacccacagg agataatatt ttgaaaagag cacacattta accagggaag
86281  aagagcccat gccctggctt aagagcatag ctgtttcctc ctccatctgc ccccattttc
86341  ccaggcctga ctagagcaag tagttaagcc aacccaaatc cagaagtatg aataatgcgt
86401  gaagcacagg aggacaggca gatggccttt tcacctgcgc aggctatata gctctgtgcg
86461  gaggggctga gagtccagcg cggcccctgt gcccctcagt aactgtgcaa gaggaagtgg
86521  gcctcccctg cgggggtatg aaatgtgggg cgaggactta ggagttcacc gggaagtgcc
86581  agccatcctg gggcagacac caagagctgg agtgatttgt ttacccTttt ttcatgagtt
86641  aatttgcttc cagcagattc cattagtgac ctgacagacc aagggcctgg aggaggtccc
86701  tggaagagct gctgaatcag cagactaaag gggcgaccag gaaggccaga cgtcttgcac
86761  cgagcccttt gctcagagtt tctatttgaa gagcttagaa taactatttg ctaagctctt
86821  gtcaaggcag cacctctaaa ccatgtatgg agagaagaca cacaaacctt gaccaccaaa
86881  tgtgaatggc caatcagaag ctgacacaat cagctgctga ctgaggaaaa atgattttat
86941  gggggaaaag aattcccaca tgaaataata agcaaaagtt ataaacttat taaaaagctt
87001  gttaaaaact tacttaaaaa gcaaaagttt aaacatggtt ccatctgcaa ctaggaactg
87061  gtcggggagg gagggtttga acttcctttg agtttcaatt gataatcccc gaagattaag
87121  aaaatagtga tactattact actgctgctg ctattattac aaataataat aactacaatt
87181  aacatttatt tggcacctag tacatgctag gaagtaggca atgagttttt ttgtttgttt
87241  gtttgtttgt ttttgagatg gagtctcact ctgtcaccag gttggagtgc aggggcgcaa
87301  tctcggctca ctgcaacctc cgcctcctgg gttcaagcaa ttctcctgcc tcagcctccc
87361  gagcagctgg gactacaggc acacgccacc atgcccagct aatttttttg tatttttagt
87421  agagacggcg tttcaccatg ttggccagga tggtctcaat ctcttgaccc tgtgatccgc
87481  ccacctcagc ctctcaaagt gctgggatta caggcgtgag ccaccacgcc tggcctgcaa
87541  ggagttcgtt atgcatgatt tcatttactc cttgcaacat ctctttgagg tagggaaact
87601  gtgtgggtga gaaaagtcac atagctgcta agtcatagaa aacgttcaga gcaaaatcaa
87661  acccaggtgt aggccaggtg cagtggctca tgcctgtaat cccaggactt tgggaggctc
87721  aggcaggcgg attacttgag gccaggagtt caagactagc ctgggcaaca tggcgaaacc
87781  ccatctctac caaaaatatg aaacaaatta gccaggtatg gtggcataca catgtggtcc
87841  catctactca ggaggctaag gtgggacgat tgcttgagcc tgggaggcgg aggttgcagt
87901  gagccaatat cacaccactg cactccagcc tgggtgacag agttaagact ctatctcaaa
87961  aaaaaaaaaa aaaacccagg tgtgtccaac tccaggcttg tgcacttacc cactctgcta
88021  tatggctaaa tcccacaggc tacagagaga gcagggaagt aggactcagg agaactgcct
88081  ccagccccac aggcaccctg actagctgtg ggatcttagc aagtgatttc aacttgtggg
88141  ccatcgtatt ccatcagatg attgatgtct tagggacccc ttcaattcaa accacagtta
```

FIGURE 3-X

```
88201   ttggtggact gcctgagtgc caggcactaa gatagggggt gggggaatgt ataaacaaaa
88261   ctaaggcgtg gtccttccct ccaggagttt acagtctagc cctgccttat gcaaaagcca
88321   aagagcctta ggtgtccctt caggtctaaa atactactca gagattcctt ttgaacatgc
88381   aaagtatttc tgacagcatt catattcatc ttctatttgt tctgtgtgga catttgcatg
88441   gagaggtatc tgaaattatg ttcgtccaat ggatttgggg tgataatttt gctctcatct
88501   ttttactttt atgtgttgct tgaattttta gtaacaaata tacatcattt ctataaaaac
88561   aaaggcattt ttaaaatact gctaaataac actatatatg ggtgtatatg tatatgtata
88621   tatgcacaca catacctata tacacatata tatttaaaca gtaactggat atttaataat
88681   actgtggaat tattgttaac tcagataaga ttgtgactag gtttccagaa agaatcttta
88741   ttttttgcag atacatactg aaataattta cagataaaat gatgtctgag atttatttca
88801   aagtaatctc tagtgtgaga gggtaggagt aggctataga tgagacaacg ctggctctga
88861   gttgcaagtt gtggaagcaa tttgggtgct aggtgatggg tataagagtg ttcacgatac
88921   tacataggtt tgaaattttc catactatac agtaaaaaag gaactccttt caaggccaac
88981   agggaagcaa catttcccca aagctaagct aagcacccag ctggcagcat ttctacctgg
89041   ggtttctacc acccttggct tcctctcttc tttgactatc tatcacagtt taacttcatt
89101   ggttctctgc agcttcctcc ccccccccc ccacaccatt ccccctaac cttctggaat
89161   cacgtgttgt atctatttgt acattccagt gcccaggcca tccctcttcg accttaccct
89221   ggcccagggg acagagcggg tgggcctgag caatgctccc acacacctcc gctcaaagtc
89281   agctgtttgc tgtagaggta gaacagcttg actactgggg gcgagggtca cagttcactc
89341   attcccccag caaatacgga gccagatgct ggggactcag cagtgaacaa gaccaaattg
89401   ctaccttcac agagttttca tgctagagaa ggagccaact ataaatgtta tctcccatcc
89461   ctcccagatt tagggccagg ggtatattca ccaagacatt cctgaatcac ctgggatctt
89521   gtcccaaata atttaaacct gaaagacatg atctagtgcc aagagctaga ataccaccca
89581   actacccatc ccaatccagt tccccattta actgacaaaa aaaaaaaaaa acagaaatgg
89641   tgtaagccca ccatgtaagt aaacataagt aaacagtggc agaggaaagg gtacccagac
89701   ttcactgtgg caggcccacc tgggaccaca ttcagaattt ttacaagctt Wgctgcgatg
89761   ttgaccaaag ttctccccac tttttttcat ttgcttgaat ttcactttta tttttattaa
89821   catcttttgt tttgttttgt tttgtttttg agacagggtc tcaccctgtt gcccaggcta
89881   gaatgcagtg gcacaatcat
```

FIGURE 4-A

>4:68275001-68368000

```
1       caaagaaata aaagtaaaaa aaaaattaaa aatgactcaa taaaagagaa atcacaatag
61      aaaattataa agtataaaaa aagagccatc acacttcatt cttcctaagt cctacaccag
121     agttgcttta aatgtgtgta tttaatggtt aaataagacc tgaaaaggga gttggcctaa
181     aaatattttt accgtRaaac atcaatttct tcagagttaa tgagaattat atgctaacca
241     taccaaaaac cttgattaga atttttagct tagatgtatt gattgatatt aaaaagggaa
301     agttgaggcc acatacagaa agcaatggcg aaggacactg cctccaacag aaaatatgaa
361     ccaaaaagca taatggaaaa acatttacat gtaatatact aaatagatgt catgccaaca
421     cagacacaca aggtgagaat accatgtgat aatgcaagta gattgaaatg ctaccggtgc
481     aagccaagga atgccagaaa ttgccagtaa accaccagaa gttaggaata aggaaggatt
541     ctctccctta caggttttca aagacatcac ggccttgaca acaccttgat ttcagacttc
601     tagcctctag aactgagata ataaatttct gttgctttaa gtcactcatt ttgtggtatt
661     ttgttacagc agttctagga aattaataca aacatgcctt ttaccacagt tctctcccat
721     tggattattt aataataaat gatttaaaat tagtatttgg aaaagatgtt ttttaatgag
781     tagacatatt aatcaggtct cttttgacac agaaatgttc tatatttcat tttcatattc
841     ttgtcaaatt ataaaaataa tatcataaat atggacatcg ctacaagcta ttatgtacag
901     tcactgaaaa atcagaaaaa acatgtgagt tagaaatgtt tatcaaatac aagtctacag
961     aaaactgaat aaaatttttа ttttaaaaca tcacaagtaa ttacaaagac aaatattcca
1021    actatacaaa ctgtttttca catatacata tatacataca tatatttaag gttggttgat
1081    ctttcatctt tttatctgag taaaaagaag aacactttcc tcattccgga gatttttgat
1141    gttataatag ttacctaaaa gtaaagtgat aagaaaataa aaattattta catatgaatc
1201    attctttttg attaagtatg tagtcatcac ctgacagttg tagtatagga cacaagataa
1261    aaacaattca cctaacctaa agtaacttat ttaattccat gaactacatc aacagctaac
1321    catgacacgt gatccagatg aggcttgaga gaacagaaat ataaatttca catcgttatg
1381    aaaataaata cggagatgaa tcagtcgaga gtgtaagaaa agaaaatctc tttctgaaac
1441    caatctttat taggtctaca gaactgctgt ggattcctct aaccagacat actcatatat
1501    agatgaatat aaaaggtaca atgttaagta gaataaaatc tcaaaatttt atttatttac
1561    ttatttttтg tttttgtttt ttctttttct ggagaacggg gtctcgctat attgcccagg
1621    caggtctcga actcctgggc tcaagctatc ctcccgcctc ttgcctccct gagagctggg
1681    attacaggca tgagccaccg cgcccggcca aaatctcaaa attttaaaaa ggcaaatgct
1741    actcttaaat aaatgaggta acaaataaca gcaaaagtga aatacagatg gaccgcaagt
1801    atcacatgtg acaggcttat tagaattgag tactatcact gtgactttct atatattaga
1861    aattgagaaa gtattcacaa cagtgctgac atactaagga ttatgcaagt atggctatta
1921    ttattgtctc cttccaacag aatgtaagct ctataaaagc agggattttt gcctgttttt
1981    tcatttatat atcccaaatc ttgcagcaat acatggacat agtagacact ttgttttgtt
2041    tttgtttttg agacagagtc ttgctctgtc acccaggctg gagtgcagta gtacaatcac
2101    agctcaccgc aactttgaac tcctgggctc atgcagtcct cctgtcccag cctcccgaga
2161    gctaggacta taggtgtgta tcaccgtgcc tggttaactt ttgcattttt tgtagacaca
2221    gggtcttcct aggttgccca ggctggtccc gaactcctca cctcaagccc tcctcagcct
2281    ctcaaaatgt tgggattaca ggtgtgagcc actgcacctg gcccatagta gtctcttaat
2341    catacttgtt gaatgagtga atgaatgaca agtggattaa aatgcctatt tacaataact
2401    tctgctgttc gatagctgat gaagcatacc attcttgaaa cagtttaaaa taatgttcta
2461    aaaacataat ttcttagaga taacatgggg taaataattt ttgttctcct ctgtaataca
2521    acatatctgc tgactatata cataaatgta ttttttttтt tgagaccgag tctcactctg
2581    tcccctaggc tggagtgcag tggtgcagtc tctggtcact gcaagctcca cctcccaggt
2641    tcatgccatt ctcctgcctc agcctccaga gtagctggga ccacaggtgt ccgccaccac
2701    gcccagctaa ttttttttgta tttttagtag agatggggtt tcactgtgtt agccaggatg
2761    gtctcgatct cctgtccttg tgatttgccc actttggcct cctaaagtgc tggtattaca
2821    ggcatgagcc accatgccca gccaatattt tcatattcag tagtaaacaa ctcttgttct
2881    catgtatcaa aacactaatt ggttgggcac tgtggctcac acctgtaatc atagcacttt
2941    ggaggccaag gtgggcagat cacttgaggt caggtgatct gttacggcta acatagcaaa
3001    atcccgcctc tactaaaaat acaaaaagta gccggatgtg gtggttcgtg cctgtagtcc
3061    cagctactcg agagtctgag gcacaagaat cgcttgaacc caggaggtgg aggctacagt
3121    gagctgagat cgtgccactg cactccagcc tgggtgacag agcaagactc tgtctcaaaa
3181    acaaaaaaaa accaccacaa aatgctaata cacttaccta atgatcaaat gaaaggaaaa
3241    tatgtaccaa actgttctga catccatcag tcatatcaca cttcattttg accactgtgg
3301    cttccatttc catagtgccc aaaaacaact atatacatgt tgtctaagtg aaataaattg
3361    tcttacctga aggaacatag aacgaatccc cagtacttaa tatataaggt gtttcatgta
3421    aagtacacaa aaggtcacca aagttaacat aaaaaacctg taaaaataaa aataaaaatc
3481    tcattcaaaa ctgaaccatg ataatacttc aaaggaacct taaaatgtag tgggaagaat
3541    aacgactccc cacaaatgtg ctggctttaa tctccagaac ttgtaaatac attaaatttc
3601    atggcaaaag ggaccttgag atggggagat tatctttgat aatgtaggta ggcccaatct
3661    aatcacatta agcccttaaa aatagagtac tttctcctgc tggaagcaga gaagaaacga
```

FIGURE 4-B

3721 aagagaagga acagtcacag aaatttgaag catgaaaggg acttgcaccg ttgctggcac
3781 taaagatgtg gggccatatg ccaagaaaca tgagttattt ctaaaatcta agaatgacac
3841 cctggccaac agccagtgag gaaatgggaa cctcagtccc acaaccactt aaaactgagt
3901 tctgccaaca gtctgaacga gactggaagc agactggaag cagattcgtc cttacagcct
3961 ccagaaagaa atacaaacct gtcaacatct tgattccagt cttgtaatac tctaaagaga
4021 atacctggtc aagctcatgg ttctctctac ataactgtga gatagataat acaagagtat
4081 tgtttcttgt tttttgagat ggagtctcac tctgtcatcc aggctggagt acagtgaggc
4141 gatctcggct cactgcaagc tccgcctcca gggttcatgc cattctccag cctcagcctc
4201 ctgagtagct gggactacag gcgcccgcca tcgcgcctgg ctaatttttt tttgtatttt
4261 tagtagagac ggggtttcac catggtctcg atctcctgac ctggtgatct gcccgcctcg
4321 gcctcccaaa gtgctgggat tacaggtgtg agccaccgcg cccagcccaa gagtattaag
4381 ccattaaatt tgtgatacag tcatgcactg cctaacgatg tttcagtcaa cagcaaactg
4441 cctacatgat agtggtccta taaggttaca atggcattta aaaaatcgta ttgcctagtg
4501 acctcacagc catcatgatg tcccagtgaa aagcattact cacatgtttg tggtgatgct
4561 gatgtaaata aatctactga actgccagtc acataaaagt atagcacata caggaaaaag
4621 tggaggatag aaagtaggac taacttgcag ctcccactca gatggacaga acagcatgtg
4681 gaaactcaca tcatgaactt ttttgccaga agaactactg caggaacata ccaagaaaac
4741 tgaaagaatt cacagactct ttgaaagaaa tggcttgctg ctgcaaactc catgagacag
4801 ctgaacaacc gtgagtgccc aaagtgtgaa agggggggaaa gtctgcctcc aaacacatcc
4861 ttactgggga agctgaaaat ccaggtcatg agagaaggat ttaaccttac ctagagctga
4921 aacgaattga gagagccaag ggaaatataa tagtagaagc agaggcagga agagccctgt
4981 taagtactcc tggtttccaa ggaaacccaa ggaagccatt tctgacttta tctcataggg
5041 ttccttgggg atggctgcca gtggaactgg gggaggacca cagaaagaag gaaacttcca
5101 gctgaacttt aaataatttt gatgaagcgt gaattttcct gggcagaatt gggggaaggg
5161 cgaataggaa gttcagatac aagccagggt aggcagcaag gggcagggcc tgaaagccca
5221 gcttgctttc tcagcaggga gacttatagc ctagcgcaaa atgtcagtcc tgctcactgg
5281 ctgtctggat acaaacttgg tttggtgggg cacagtagga gtgagactgg cctagcctgg
5341 ctgcttggga gctgggtgag gccagtctgc cagcttcccc cacttccttg gtgaccagta
5401 tgatgcacta gagacagcca taatcccctt gggaacataa ctccagtggc ctgggaacca
5461 tattttcatc ccctacagtg gtcacaaaaa gctcagccca aggagagtct gagctcagac
5521 acacctaatc aatcctacct gcacctgatg gtctttctct aactgccctg tagcctaaga
5581 caggagctat aaggccccac ccatcacctg agaaacctga atacttaccc aggcaacctg
5641 tggcaacctt gtatcagcag atgctctctt gaaagtacca tctcctggct ggtggccagc
5701 cacctgctag cacaaccaat attaaagaaa accagcacac taaacaaaac tacaaccaaa
5761 gaccgtcaca gagtctgctt cactcccctg ctacctccac tggagcagct gctgagatcc
5821 atggctgaga gacctgaaga cagatcatat cacaagacgc tttgcaggca cttcccagta
5881 ccagcctgga gccccatagc tctgctgggt agctagacac agaagagcaa tagcaatcac
5941 tgcagactgc ctctcaggaa gcataatccc taggggaagg aggagagcac cacatcaagg
6001 gatcactcca tggaacaaat gaatctgaac tgcagtgctt gacctccaga tgtctcctct
6061 gacacagtct acccaaatga gaagaaacca gaaaaacaat tctggtaatg tgataaagca
6121 aggttcttta acacccccaa aagatcatac tcgctcccta gcaatagatc caaacaaata
6181 aatctctgaa ttgccagaaa aataattcag aaggttgatt attacattac tcaaggaggc
6241 accagagaaa gatgaaaacc aacttaaaga attttttttt aaaaatacag gattttgatg
6301 aaaaaatctc gagagaaata gatagcataa ataaaaaaca atcacaactt ccagaaatga
6361 aaggcatact tagagaaatg caaaatacat tagaaagttt taacaataga atcgaacgag
6421 tagaagaact tcagagctca aaaacaaggc ttttgaatta acccaatcaa agacaaagag
6481 aaaataaatt aaaaaaaaaa atgaacaaag ccttgaagaa gtttgggatt atgttaaacg
6541 accacacata agaataattg gtgttcctga ggaagaagag aaatctaaaa gtttggaaaa
6601 atttgaggaa ataatcaagg ataacttccc tggccttgct agcaatctag acatccaaat
6661 acaagaagtt caaagagcac ctggaaaatt catcgcaaaa agatcatcac ctagccacac
6721 agtcattagg ttatctaaag tcaagacaaa gcaaagaatc ttaagagctg tgaggccaaa
6781 gcatcaggca acctatttaa aaaaaaccta tcagattaac agcagatttc tcagcagaaa
6841 ccctacaagc tagatgggac tgaggtccta tctttatcct ccttaaataa aacaattatc
6901 agccaagaat tttagtatcc agcaaaacta agcttcataa atgaaagaaa gataaagtct
6961 ttttcagaca aacaaatggc gagagaaatc cccactaaca agccagcact acaacaactg
7021 ctaaaaggag ttctaaatct tgaaacaaaa actcaaaata tatcaaaata gaacctcctt
7081 aaaggataaa tctcccaggg tctataaaac acacacacac acacacacac acacacacac
7141 acacacccaa agtattcagg caacaactag cacaatgaac agaacagtac ctcacatctc
7201 aatactaaca ttgaatataa atgacttcaa tgctccactt aaaagataca gaacggcaga
7261 atgcataagc attcaccaac aaagtacctg ctgtcctcaa aagactcact taacacataa
7321 ggattcacat aagcttaggg taaagaagta gtaaaagata ttccatgcaa atggatacca
7381 aaagagagca ggagtagcta ttcttatatc agacaaaaca gactttaaag caacagcagt
7441 taaaaagaca aagaaggcca ttatgtaata agaaatggac tagtccaaca ggaaaatatc
7501 acaatcctaa atatatatgc acctaacacc agagctccca aatttataag acaattacta

FIGURE 4-C

```
7561   ctaggcctaa gaaacgagat agacggcaac acacttatac tcaaggtctt caatacccca
7621   ctgacagcag taaacaggtc atcaagacag aaagtcaaca aagaaacaat ggatttaaac
7681   tatacactgg aacaaatgga cttaatagat atttacagaa cattctaccc aaaaactgca
7741   gaatatacat tctattcatc agctcatgga acattttcca agatagactg tatgagaggc
7801   cacaaatcaa gtctcaataa atttaagaaa accaaaatta tatcaagtac tctctcagac
7861   cacagtcgaa taaaattgga agtcaactcc aaaatgaacc ctcaaaacca agcaaataga
7921   tggaaattaa ataatctcct cctgaatgat tgttgggtca acaatgaaat aaagatgaaa
7981   attgaaaaat tctttgaact gaacaataat agtgacacaa tctatcaaaa cctctaagat
8041   acagcaaaag cagtgctaac aggaaagtta atagcattaa atgcctccat caaaaagtct
8101   gaaagaacac aaatacacaa tctaaggtca cacctcaagg agctagagaa ataaaaacaa
8161   acaaaaccta aacccagcag atgaagatca ggccatttgg tagaactaaa tgaaaatata
8221   atcagaattg atagaccatt agtgaggtta acaaagaaga cagaagatcc aaataaactg
8281   aattagaaac aaaatgggag atattacaac caataccaca gaaagacaaa agatcattca
8341   aggctcgtat gaacaccttt atgtatacaa accagaccta gatgagacag ataaattcct
8401   ggaaatatac aactccctag attaaaccag gaagaaatag aaactctgaa aagaccaata
8461   acaagcaatg agattgaaac ggtaataaaa gaattgtcaa caacaaaaaa agtccaggcc
8521   gggtatagtg gctgatgcct gtaatcccaa cactttggga ggctgaggca ggtggattgc
8581   ctggggtcag gagttcaaaa ccaggctggc caacatggca aaaccccgtc tctactaaaa
8641   atacaaaatt agccagaatt ggtggcgcat gcctgtaatc ccagctactt gggaggctga
8701   ggcaggagaa tcacttgaac ctgggagatg gaggttacag tgagctgaaa tcgcaccatt
8761   gcactccagc ctgggcaaca agagcaaaac cccgtctcaa aaaggaagaa aagaaaaaga
8821   aaaaaaaaat tccaggacca gatggattca cagctgaatt ctatcagaca ttcaatttat
8881   ctatcagaaa agagacaaat tcttggaaat acacaacctt cttagattaa accaagaaga
8941   aatagaaact ctaaatagac caataataag cagcaagatt gaaaaggtaa taaaagaatt
9001   gtcaacaaca acaaaaaagt ccaggaccag atggattcac agctgaattc tatcagacat
9061   tcaaagaaga attggtacca atccaactga aatgattcca aaagacagag aggagtcaat
9121   acgcaagtca ataaatgtga aataccacat aaacagaatt aaaaccaaaa aatcacatga
9181   tcagctcaat agatgcagaa aaagcatttg ccaaaatccg gcattgcttt atgattaaaa
9241   ccctcagcaa aatcggcata gaaggagcat aactcaaggt aataaaagcc atttatgact
9301   aacctacagc cagcatcata ctgaacaagg aaaagttgaa aagtttctcc tgagaactga
9361   aagaggacaa ggatgcccac tttcaccact tctattcaac atagtactgg aagtcctggc
9421   cagagcaatc ggacaaagga aagaaataaa gggtatccaa actagaaaag aggaagtcaa
9481   actgttgctg tttgccaatg atatgatcat atagctagaa aaccctagac tcatccaaaa
9541   agttcctaga tctgataagt gaattcagta aactttcagg atacaaaatc aacgtacaca
9601   aatcaatagc tctgctatac accagcagcg accaagctga gaaacaaatc aagaactcaa
9661   ctccttttac aatagccaca aataataata ataatttgga atatacctaa ccaaagagat
9721   gagaaatcat agacgacata aacaaatgaa aacacatccc atgctcatgg atgggtagaa
9781   ccagtattgt gaaaatgaca tactgccaaa agcaatctac aaattcaatg caattcccat
9841   caaaatacta tcatcattct tcaaagaact agaaaaagca atcctaaaat tcattatgga
9901   accaaaaaag ggcctgcata gccaaaggga agactaagca aaaagaacaa atctggaggc
9961   atcacattac ctgacttcaa actatactat aaacatatag ttaccaaaac agcacagtac
10021  tgaaatgaac ataggcacat agaccatgga actgaataga gcctagaagt aaagccaaat
10081  acttacagcc aactgatctt caacaaagca aacaaaaaca ctaagtaggg aagggacacc
10141  ctattcaaca aatggtgctg ggataaccag taaggcacaa gtagaaaaat aaaactggat
10201  cctcatctgc caccttatac caaaatcaac tcaagatgga tcaaagactt gaaactaaga
10261  cctgaagcca tgaaaattct aaaagataac atcagaaaaa cccttccaga cattggctta
10321  ggcaaagact tcatgaccaa gaatccaaaa gcaaacacaa caaaaacaaa atagatggga
10381  cctaattaaa ctaaaaagct tctgcacagc caaagaaata attggcagag taaacaggca
10441  acggcagagt aaacagacaa cccagagtga aagaaaatat tcacaaactg tctgtgcaac
10501  aatggaataa tatccagaat ctacatggaa ctcaaacaaa tcagcaagat aaaaccaaac
10561  aatctcatca aaaagtgggc taaggacatg aatagacaat tctcaaaaga agatatacaa
10621  atggccaaga agcacacaaa aaaaatgctc aacaataact aatgataagg aaaatgcaaa
10681  tcgaaaccac aatgagatac cacctcattc ctgcaagaat ggccataatc aacaaatagt
10741  agatgttggc ctagatgcag tgaaaaggga acaattttac actaatggtg ggaatgtaaa
10801  ctagtacagc cgctacagaa aaccttagta gagctaaaag tagatctacc atttgatcca
10861  gcaatcccac tactgggtac ctactcagaa gaaaagtgag tcattatatg aaaaagacac
10921  ttgcacacgt acatttatag cagcataatg tgcaactgca gaaatacgga accaacccaa
10981  atgcctatta atcaacaagt agataatgtg gtatacatgt accatagaat agtactcaac
11041  cataaaaagg aatgaaataa tggatggtat ttgcagcaac ctgcatggag ttggagacca
11101  tcattctaat aaagtaaccc aggaatggaa aaccaaacat tgtatgttct cacttataag
11161  tgggagctaa gctatgagga tgcaaaggca taagaatgat acaatggact ttgaggactc
11221  agggggaagg gtgggaggag gagaggtata aaagactaca cggtgggtac agtgttcact
11281  gttcaggtga taggtgcacc aaaatctcag aaatcaccac taaataaata atctcttgta
11341  atatggtttg gctttgtctc cccacccaaa tgtcaccttg aattgtagct gccataatcc
```

FIGURE 4-D

11401 ccacatgtca tgggagggac ctagtgggag gtagttgaat catgggggca gttaccctca
11461 tgctgttctc atgagagtga gttctcacaa gatctgatgg ttttataagg ggcttttccc
11521 cctcgcttgg caattctcct tgctgccacc atgtgaaaaa ggacatgttt gcttcccctt
11581 ctgccataat tgtaagtttc ctgaggcacc cccagcccca cagaactgtg agtcaattaa
11641 acctctttcc tttcaaaatt acccagtctt gggcagttct ttataggagt atgagaaggg
11701 actaatacat cctgtaatcc cagctacttg agaggctgag gcacgagaat tgcttgaacc
11761 cgggaggcag aggctgcagt gggctgatat cgtgccactc cagcctgggg acagtgggag
11821 actctatctc aatgaaagaa aaaaaaaaat tatctatgta accaaacact acttgttctc
11881 caaaacctat tgaagtaaaa aataaaatat aacacaattt tgtgtggtac aaaacacttg
11941 gtaatgataa taaacaacta tgccactggc ttatgtattt actatacttt tatccattac
12001 cttagagtgt actcttacca cttatataaa aaaaaaaaaa gttaactgtg aaacagtctc
12061 aggcaggtcc ctcagtaagt attccagaag aaaccaatgt tattatagga gatgacagct
12121 ccatgtgtgt tactattccc aaagacctta cagtgggaca agatgtggta gtgcaagaca
12181 gtgataatga tcctaatctt gtgtagccca aggctgatgt gctgtttgtg tcttactttt
12241 taacaaaaaa gttttaaaag ttaaaaaaaa aatagaaaaa agcttatага ataaggaaat
12301 aaaatatttt tgtacagcca aacagtatgg ttgaatttta agtattacaa aagagtccaa
12361 aagttaacaa caacaacaaa aggatataaa gtgaaaatgt tacagttaga tgaggtttat
12421 tactgaagaa agaaaaattc tgtttatgaa tttagtgtgg cctaagtcta cagtatttat
12481 aaagtctaca gtagtgtaca gtaatgtcct aggtcttcac atttactcag cacttactca
12541 ctgactcatc caggaccact tccagtcctg aaagctccat tcatggtaag tgccctataa
12601 gggtaacatt tatcttttat gccacatttt tattgtactt tttctacatt tagatacaca
12661 aatacaactg tctacagtat tcagcacaga aacacgctgt acaggtttga agtacagaag
12721 caacagggta tcatcctata tagcctaggt tcgtagtagg ctatatcatc ttatgacgtt
12781 cataagacaa aatagcttaa tgactcattt ctcagtaaca catagctgta atttgttaga
12841 gcagcaactg aaaactaata cagatgtccc caacccccag gacagggacc gctactggca
12901 gtggcctgtt aggaatgggc tcacacagca ggaggtaagc agcacgccag tgagtgttac
12961 tgcctgagct ctgcctcctg ttagatcagc agcagcatta gattctctca ggcacatgaa
13021 ccctattgtg aaatgtgcat gtgagaaatc taggttgtgt gctccttatg aaaatctaat
13081 gccaactccc actgccacca ccaatttttg tggaagaaaa attgtctttc acaaaccagt
13141 ccttggtgcc aaaaaggatg gggaccattg taatacagga agagatttct tttatttttc
13201 cttattttgt ttttccttat tacttactgg cttatgaaaa ttctacagcc actgtaaaga
13261 ttatcaccta cactataaag attatcaccg cactgggcgc agtggctcac Rcctgtaatc
13321 ccagcacttt gggaggccta ggtgggcgga tcacgaggtc aggagatcga gaccatcctg
13381 gctaacacag tgaaacccca tctctatgaa aaatacaaaa aattagccag gcgtggtggc
13441 gggcacctat agtcccagtt actcgggagg ctgaggcagg agaatggcgt gaacctggga
13501 ggtggagctt gcagtgagct gaggtcacgc cactgcactc cagcctgggt gacagagcga
13561 gactccaact caaaaagatt atcacctaca ctaaaaaaat acctaggctt tatgtatgcc
13621 tctttaaagg tcacatgctt aggacaacaa actactcaat ctgattgaag aatcaaagaa
13681 agcaaaaagc taaatcccta ttcccctcaa aatcatatgt taaacaagca tcaagactaa
13741 gaaagtaaaa gaggataaat ctctcaatca tgctggcttt catattcaaa catgccaaat
13801 ttcatacttt atacataggt agcacacact ggcagtgacc tggttttttg agcaactaaa
13861 ataaacagca taaggccttc aaactttgag attcagccat tgttagaaat tagtttgtct
13921 cctgaaaatt cacactctgg aatcttaacc cccccaatgt gatgggatca ggaggtgggg
13981 aatttgggac gtaattaggt cctgagtgca gagccttcag gaatgggact agtgccctta
14041 taagagacac agatcaaagt ctatggtatt cttattatag cagcccaatc tgactaagac
14101 atccatcaag aatgttttgc ctcagataat tatttctcct ccttatttтt tattactact
14161 ttgtacaatg atcataatat accagaataa caataacagc tagcacttat atagcactta
14221 ctctatacca agcagtatta tgagaacttt attcttattc acttaatgta tatgcccact
14281 aatgagacag cgactacttt agtagctttt tatggatgaa gaaactgagg cacagagagg
14341 ttaagtaatt ttccaaggtc acaatgtaaa tggcaaaggg ggggagtttg aactcaggca
14401 tttagattac agaattcatt ccctaatcac tgtgatatac tgcctcccta aatggggaaa
14461 aaagcccccc gaaatgtttc aaaggYattt ttataactta ataagacaaa aaaagttatt
14521 acatccaaaa ataatttgcc ttaatcaaaa agacaaaaaa gtaatactaa aaagtatata
14581 tttctagtgt ttttttctca tctattcaat ttggttctct aagcacacct ctcctcattt
14641 ctgtaaacat ttaattaaga aagcttttga aacatacaaa aatagaatga Wgaactctca
14701 tgaatactta atgctaacca cctgcccaag ccagagtcta ctatcctcac tccatcttat
14761 ttcaaaataa atgcccagta tcataattca tctgcaaata tttcagtatg tactgaaaga
14821 tcaagtgctt ctaaaaaaca taaccgcaat actgttatca cacctaaaaa ataaatattt
14881 aataccaaat atctagttaa tagtcaaatt tccaatcatc agttaaacat atcttcaagg
14941 atttctttga atcaggatcc aggttagcac cacatacagt gattggttga tctgtttctt
15001 aaatttctca atcaatatta gcggttctca atcaaggtga ttttgcctct cagggacaac
15061 ggacaatttc tgaagacata tttgattttg tcacaactac agtggtgcta ctggaatcta
15121 atgagtagag gccaggtaag cttctactta cagggcctta cctacaggac ggcccccact
15181 acaacaaaga agtatctaat caaaaatgtt actagtgcca tggttgagaa accctgattc

FIGURE 4-E

```
15241  cctaggtttc acctggttct tttttatttc ttacatttaa tttattgaaa acattaagca
15301  gtttgtcata aaaatcttca tgcaggtgga ttttgctgac tacatcccca ggtatcattt
15361  gaaagatttc accacacttt atttcttgta aactggtagt taaatctaga gacttgatca
15421  gattcaggtt ttgtatttgc ttttaggttt ttttttttct ttttcttttt taaggcatgg
15481  ggttggcaag actacttctt ggatagtgtt gtatttttct atcaggaggc acataatgct
15541  tggttgtatt Stttttagtt atctgtcagc cactgagcct tgatgcctag atcccaatag
15601  cttatcaggg attgcaaatg gtagtatttt aactccaaca ttccttcttc actcattagc
15661  tggaatatta ttctaataag aacaaactca tctattattt gattacctaa tagcacagtt
15721  tatatagaaa aggcagaatc agttgttaaa gtgagttggc ttcctagaat cctccaacag
15781  taacaaattt ttgttgtagt agtagtagta gtagtcttag aaattcatgg atttagatat
15841  aattgatata tatcaatcta ctgaagttat tcttgttaat actcaagttt catatcctta
15901  gccagccagt ggtagcatct tcaagttgga acctgaaaca tgacccaatc ttaatggttt
15961  ctgattagta gtttgctatc gattccaagc tcagtttcag gcctagaatc aggcattcct
16021  ccaaggaact ggttctttta agtaggttgc tgtcaccttt ttgtttttct agctctgaaa
16081  catttgtcta attttaaaaa tattctcaac cattttcat tgggtttttt ctctctcctg
16141  aaattcttat tatttgtatt taggcatgct tttctctctt ctctacatta tttaacttta
16201  atattttcta ctttttaatt tccttcttgt ctcctgggaa gttctacaac cagctctcat
16261  aattaatatg taaattcttc tgtagtttga ttttaataaa cctactactc ttctctttta
16321  aatattctct agcagtgttt gaaagagagt taatggccta tgctggttct ccatcttgtc
16381  ctgtttatta tctttttcaa aagcatatat ttgtaagtgt gcttcatgtt ttaaaatgta
16441  aaaatacgag tcttattaaa aaaaacttca gctttaattt tcatttattt tcattataaa
16501  aagttaaagg taatatatag aagaaataaa atcaactaaa ataactacat aatcacagag
16561  taaaacatgt attctcttca agattttatt tacattttta aattctagtt agaaaaatag
16621  atactatata caaagaactt ttatcttgca tatttgctta aacatttatc ccaagggtct
16681  tcatatatgg ctgttcttca aaataaaatg ctcaaactaa ttttaatgtg cttagtgcag
16741  tacctggaac acataagttc tgtataactg tttattatat cacagaatgg tttatcccat
16801  tactaagtca aaattaaaat ataaccattt tcttcctgta gcacattgta gtggtttcta
16861  attctgcaat taccatcaat gattcaataa ataactacat atgttcagca cccctcactc
16921  tgaacttcca aattacttcc ctgggaaaaa tttctagaaa tgagtcacag gttacagggt
16981  atgaatactt ttctaataat ttccattaac tcagtaacta taaagtgtca tttggattta
17041  aatgtgtatt cctttgaata ttagccagat tacacacttt taaacatctg ttactcaaat
17101  ttccttttgt ataataaagg cattacagtt taataggtaa ctgctattac ttcagtgatg
17161  ataattcatg tttatcctca ttattagtta ctttattatt atacagggac catctatgct
17221  tttaaaatgt gttattttca ctcataataa taactggagg cctataacat tttcagtttg
17281  aattaatgaa tattgttaaa ttatttaaa tcatgattct aactatactg atgtagtaat
17341  aatctttgac tttcatatgt atcttttcct tgccttgcct aaatttgccc ttctttgtta
17401  atataaactc aagagtcttc acaatttttc cctccttttt gaatttactt taagtttgtt
17461  ggcaaagaga atcattctga ctgtgatact ttagtagcac aatttcagac tcctgaaatg
17521  tcagatactg aaagatgttc cttagctata ttttttaact gctggtataa tgttaaatat
17581  tcttaaaatt tggtgctttt caaattctat attctacaat tttcaaactc agactacaaa
17641  tatttttat atatataaaa tattcctgtt atacatacac acacacacac acacacacac
17701  aYacacacat ataaaacaga ccaatgttgc agaggtgtat gtctctaaat gggggtgggg
17761  gtggaagaaa ggagataaag agaaagaagg agaagataga aactgactca ccttgaatct
17821  cattacaact tcagccagta atattgttac atataatagt atattattta gYacttactg
17881  aagcaactag cttagtcctg agtctatgtc taaatagtta agtagttttt ttataatctc
17941  actcatttta tttacttttt tctcaggtta ctgttttatt gatatataaa aattttacat
18001  ggtacatgtc acattttaat acctgtatgc aatgcataat gatcaaatca gggtaactgg
18061  gatgtaagtg atttttgaag aagtctttta acctatttct ctactgcatt atcatagaaa
18121  tgtggaatta gatttaatct ctataatccc ccccagcact aaaattctaa tgatttgttt
18181  gcaaacctat atttaagatg tttttaagta aagaacattc cttaaatttc tgaatctcaa
18241  aatacagtgc tatgcagaga acagatattg tcaaataact gttggctggt tgactctaaa
18301  cagatttaac aaataaaaac ctttctctgt gtggtgtagt aatcgaattt catttctgag
18361  aaaacagcat tttaatgtct catcaaatcc aataaatcag agttctacca agagtgaaac
18421  ataaatatat aaaaaaagta ctcaccaata tatcctggcc aacatgctgc tttccctttt
18481  cttcttgtgg tcctaatatc aatttcccag tagaaaaaaa gggtgtatcc aatgtcttgt
18541  ataccttcaa ctcaccatgc ttaacaaaaa attgatatgt atcttgtggc cttacaagat
18601  ctaggagtaa aacagtaata caaatataaa acaacagcag tattggttaa tgatataaca
18661  tatacacata tatataataa agagtcattt ctggatataa atctgcctta gagagttgtt
18721  tcagagcatt aactctggag ctagaatgcc agaggtcaaa tcctgtctct gtcttaaaag
18781  ctgcatgacc ttaggcaagt tacttaagct ctctctcact gtgcctcaat ttccttttcg
18841  ataaaattaa ataataataa gttcatctat tcgatattaa tgagctgctg tgaatatcaa
18901  cgagttaatg tgtgtaaact atttaaaata ttgtctggca aaaagtacgc actgttagca
18961  ccagctgaat aacaaaatgg cagtattata ggaactgaga cagaaaatag aattagctga
19021  gataaaaaaa agaattctct atttttaaag tttaacataa accacaaatg tatcatcacc
```

FIGURE 4-F

```
19081  atgccaatag attttttttc ctgcttgact aataattttt attaggtaaa gtttgtgaga
19141  aagcagagtt ctaatggttg ttcataaatt actacctcac ttacaaaaat ttccattact
19201  taacaaactt ttccagatct ataattaata aattattctg tcatcaatgt ctgattctgt
19261  ttcccaggat atttgggaat gttagtgatc ctacagacta catgatatag aaataaatat
19321  atatatatat atatatatat atatatatat atatatagaa atatatagaa atatatatat
19381  agaaatatat atatagtaga ttcaagaaac aaataagaaa tgaacttttg acaacactca
19441  caataacctg gtagagtaca gcaatgcttt cagttaagac actaaacaaa actccatact
19501  tctaagctct gttaccaaat aaataatctg catatagaga taccttttgct aaggtgttaa
19561  gatattaaat gaaatctatt ttcatggYca aaattttaag taggagtacc attattacta
19621  cctttctttt tttaggaatt aactaactag ttttattatt attattattg agacagggtc
19681  tggctctgca cacccaggct gaagtgcagt ggcatgatct cggctcactg caatctccac
19741  ctcccagact caagcaattc tcctgtccca gcctcccaag cagctgggac tacaagcacc
19801  caccaccacg cccagctacc attaccgcgt ttctgatttg acctacttct cctctgatat
19861  aggacaaaca gcccttagta ctcaatggta cagtcataac agttgttaaW atatttaaat
19921  atttcatata tatatatata tatatatata tatatatata gagagagaga gagagagaga
19981  gagagagaga gagagagaga gagagMgcct ggttgttaaa cagttgcgta atccataaac
20041  tcctatccca ctatagccca tcccttgggg atcttacatg acctgctaga gactaaatat
20101  ttaatccaat acagcaggag gcttccagga ctttgcagca ccagctttaa ggttaaacct
20161  ggcatctgta caactgattg atgcccttgc tagtcatcct cacaaaggat tcatatccca
20221  ggtaacatta ctgccatctg taggcccaga agtttgtctg tgccctaccc acaatccctg
20281  cctggcttca tgataagcgt atccttgagt ctgccaacta tctatctaca attccccata
20341  aacagcctct gttcccatgt gtcctcctgg tcaaagaagg cctagaccct gttgggggag
20401  acgacattcc tgattgtagc actacttgtg ccccagagga caagtccggg gtgttaagtg
20461  tcccctggag ctgtctgata tcaatattcc cccacagttg tcttcacata ggttgctcgt
20521  tggaattaga ctcttctgga ctttagcact gatagaattg gggagagaaa gggaRagtag
20581  acaggccaat catcaagcaa gcaatatttt cttacaaatt gtttaattaa gtattttcag
20641  tatcatgcct gatcttaccc atgagaataa tctctcttgt ttctgggtcc tttaccctcg
20701  ttggctgcaa aggatctcca agaggtatac ctagatttac cattaagtta gtctctgcaa
20761  aagaaatacc attgaaatac aaactactta cttaaaatta ttctcaatcc caaaaagttc
20821  agtacaaaat atagctatag agaagttata aaactgtcaa acgcaaagaa gtctcttcct
20881  atttataaat cactctgata aacaaaaatt caaacttaaa aacccataat gatttaatat
20941  aaaagcaaac accctaaact tcctatttтc tgtaagcctt aacatgtaaa ctcttgcatt
21001  atcaaggtat taagtgatgt tttgcataat ctctaagtga gatagaggta tataaaacac
21061  tttagaaatg tctatgttaa ccatagcttg gRcataggaa tggtgggttt catttgactg
21121  ggaaatttac tcatgaagaa atttataata tgctattaat aatttttaa gaataaaatt
21181  agcttacttt ccatttggaa gaaagtagac attaaaaaat gaattttacc tatggtatta
21241  tttctgaagt ttttggaagt aaatttcata ttttgcaaag agcttcttct caaagtactc
21301  ctatttaaaa taaaatctaa aagttacttt cagaaataag ttcatacttc tttcatcgtt
21361  atcaagacag atccttttct tattagattt tttgttgact tttccaatat tttcttttgc
21421  cttcctttta gacgatattg tgtctggaga tagtactcca ctaatcacga atcctcctga
21481  aatttaacaa aaaaagtaaa atatacatgg gaaagacaaa atcttgaaac aaaactcctt
21541  aaatttaata tggcacattt cctttcaaag taccaaagaa tatatgtctg cagctaatag
21601  ttaaaataac acaaatacca accaatttca ttatatctct accaccctaa atcagaggct
21661  gataaaacag gattagatga tgaaacatga ttagatggtc ctattgctcc cctgaaatca
21721  acagtcatca gtggaggctt ttaaaaaaaa aaaaaaaagg agggggctag caatgaacta
21781  tttatagaaa tccacatgag tttagtcctt tgacttagat atatccatca aagaaagaaa
21841  tatggccagg cacagtggct cacacctgta atctcagcac tttgagatgc caaagcggga
21901  ggactgcttg agcccataag atgagaccag cctggacaac atagcaaggc ctcatctcta
21961  caaaaataag aaagaaaata actctacgat taacactaca ctttcttgcc caccctacac
22021  tactttgtag aaaactatga gaaaatggaa aatggttaat aggaaaaata atgaaaacta
22081  ctttaccttt accagttaca attttctcac gttctaatct aacgctttat attatgcaat
22141  tggaaatctt gtcagatgat acaggaataa caacattcag ctaccaaaga gaaaatatca
22201  gcatctctac attaatcaga caaaaatgaa tttaaatcct attatcaaaa tttgggggggg
22261  tgatgattaa gtaaggtaat gtatataaac atgtagtgca gtgtctggca catagtaaac
22321  aaatttttta ttgacagata ttataacata aataagtttt acctgatggc cttccttgat
22381  aatctattcg ctctcctcgc cagtactcca aaggtttcaa acgtgttctc ttggtcctgc
22441  gaacatttgg tgtgttggag ggcaatacta taaaaagatg tcagacaaaa gtgtatacat
22501  agtttttagt tgatggtact tagcagtgga aagactgtct tttagaacat gaaagcttta
22561  ttttttagaa catgaatatt tttctagaca aaagactaga aaaaacaaat atttatgcta
22621  acaattaaag aaaagccaat ggaaaaacag gctaccaaat tggcaaatat ttaaaaacct
22681  aataatactc agcgttgaca atgttaaatg tacactcatc ttatctgaac aatgcaaagt
22741  tctgtatact aaattgagtt tgttggctta tcttcaaatt cccatgttct tacctatttt
22801  ttgtttgcta aagctatgct gaagctaacR agaagttgcc tactgctctt aggcataatg
22861  ctcaatgttc ccctctgtaa atactgaagc cctgctcagg aacatttttа tttccctttc
```

FIGURE 4-G

```
22921  cctactcctg attcatcatt tcaccccaag acaagcatta agtattgaga ggtatagcag
22981  tctctcatca cataggccca caaaatactg acaatatcta agttctggag taaggcaatg
23041  agagaaagct tcactcctat tggtaatctg agtatgagtg aaagagcact cacagcctca
23101  aataaaggga gaaacatgga aagccaccca acatcacaga aaggtgggaa ctatcatgta
23161  tcttgactgg caaatatctc tttgggaacc atggattgtt aatctttaca tgggcctctc
23221  catgggttcc acagttctta gctttgtcta cttactatcc tagctactgc tttgtttagg
23281  agtcaactgg attaaagtgg aaggagatgc tgtcaaggaa gccatctacc tggtagttac
23341  agtttctttg acaaaattct ccaccctagc ttcatagcta agggaggatc atgcccccta
23401  ctatggggca aaaagtgaga gggcaatagg atttataaat ggacaataga ctaaYtcaag
23461  tggcttgtgg tgacttacta tgaaattcat aaaaactaag tgcataattt taatttatct
23521  gccaattaat ccacataatc gtatttctcc tgcttaagaa gaaataagtt atatttagaa
23581  acagaaactt aaggtatttt tctcaatgaa aatattaaaa aagaaaaata ttcttaccta
23641  gtttgtgatg gattctgttc tttggtatca ctttagattg ttttgagtca tctacaaaat
23701  gcaaaagata atatcRtaag aaatgactgc taattccatg ttaaattaaa atgtgtttcc
23761  tggtcaattt cagtatgagt attttaaata aagtttgacc tgataatgta ttttattact
23821  ccttttgctt ttcattgctc aactataaaa actaaggaaa agaactattt ctctagactt
23881  tgctacaatK aatacattgc aagtttgtcc aatccacctt attttgtggt tgttctgttt
23941  tgttttgttt tagaccttta gcagcctgaa accatgcttt ttagtttctg tctctggcga
24001  Yaagtggaaa agagggatga ggaaggggtt ttattggacc aaccagaaac agaaactaag
24061  aactcatgac tgtattcgct cccttggaca cccctggttt ttattctaaa cagtgtaaca
24121  gttaaaacaa acaaacaaac aaaaaaaaca acagattttc tattgctcat ggaaagatga
24181  aataagccac ttgtggataa aaatgtaggg cctgataatg gataattaat atgctattta
24241  tgaactatct gtaactttct ctttaaagtt ctaaaatagt gtaagtgtac tgaatttagc
24301  agtgtgccaa taagctcagg ggttctcaac tttggtaact agcacctaaa gatggctgtc
24361  atcccttctt tgcctttcta tgcacaaatg ttctgtatca agaagcagag tctcgttccc
24421  ttctccttca atctctgctg gagttagtga cttgcttaac taatagtatg caacagaaat
24481  gatgttctgg ggcttaaaag gctaagtcct aatacaatct acaggttcca tctagaagtc
24541  ttgggatctc actctagggg aagacagcaa caatatgaag attaacacaa gactgccatg
24601  ctgtgaggaa acctcatgtg gccacatgga aaagccacat ggaggaaaag agatgcttgg
24661  ccaacctcaa tgtcccagtt cttccagctg ggtccaaaca tgtaagtgag gaagttacct
24721  cgaatgtcca acccatttga actttcagat gactccagcc ccagcagctt gcatttaact
24781  atataaattc atgagtgtac tcaagtggca attacccaac taagccaagt caattcatag
24841  aaccatgaaa gataaatcaa tcactgtatt aagcctttaa gttttggggt ggttggttac
24901  gcagcaatat acaactggga cacaatgtcc atgaaccacc taaaactgca aatacaattt
24961  tgtatgaaat atataccttt tttccctaaa aagcRggtct gagcttttat catattctca
25021  aatgagtcca tgaaagcaaa caaacaaaaa ttgttaagaa tgattgaatt agatatttta
25081  taaaattgct ggtttggaag aaacttcagt gatcataaca tctaatttcc ctgtcagaaa
25141  caggcaaatc tctagacatc agaagtagat tagtaattgg gtaggcccag ggaaaggaaa
25201  tgaaagtcgg ctactaaagg gtaaaggggt ttctttttgg gggttatgaa aaggttctaa
25261  catcgattgt gtcaatgact acatatcatt gatttgtaca ctctaagtgg attaattata
25321  tggtatgtga agtctagctt aataaatctg tgtacccccc accccacccc caccccacc
25381  cccaaaaatt caatttccta tgcagtctcc cctagaacaa gtagaggtct catctctgtt
25441  ggaaattctc ctgtgacact gggcctcgga ggcagtgtca gagtgataca aggaaccaaa
25501  gaaacaagta aatatactga ggataatggg agtcaggttt ctcactattt gagaagaact
25561  tgtaataaat acaggcggcc aggcgcagtg gctcacaact gtaatcccag cactttggga
25621  ggctgaggca ggtagatcag gaggtcagga gttcgagacc agccttacca acatggtgaa
25681  accccatctc tactaaaaac acaaaaatta gccgggcatg gtggcacacg cctgtaatct
25741  cagctactca ggaggttgag gcaggagaat cgcttgaacc cagaaggcgg aggttgcagt
25801  gagccgagat cacaccattg cactccactc tgggcgacag agggacactc cgtcttaaaa
25861  aaaaatttaa aaataaataa ataaaaatac aggcatacct cagagatatt gccagttcag
25921  ttccagacta ctgcagtaaa acagatactg caataaagaa gtcaaacagt tttttagttt
25981  cctggtgaat ataaaagtta tttcttttgc tgttttgttt tggtttttg agactaagtc
26041  tcactctgtc acccaggctg cagtgcagtg gtgggatcaa gactcactgt ggcctcaact
26101  tcctggcagc aaatgatcct cccgcctcag cccgccgagt agccggaacc agagatgcac
26161  accaccacgc tcagctaatt tttgtatttt ttgtagaaac aggtttctgg catgctggcc
26221  aggctggtca caaactcctg ggctcaagtg accggcctcg cctcccaaag tgttgggtgg
26281  ctcccaaagt gagccaccat acccagccaa aagttatgtc tatactatac tatagtctat
26341  taaatgtgca agagcattat gtctaaaaaa atcagtgtac atatacctta atttaaaaat
26401  attttattgc taaaaatgct aaagatcagc caggcatagt ggctcatgcc tgtaatccca
26461  agactttggg aggcccagga gggtggatca cttgaggcca ggagtttgag acaagcccga
26521  ccaacatggc gaaaccccat ctctactaaa aattaaaaaa attagctggg tgtggtggtg
26581  tacactttgt aatcccagct actcttggtg gctgcggcaa gagaatcgct tgaacccagg
26641  aggtaaaagg tttcagtgag ccgagatcac accactgcac tccagcctgg gtattagagc
26701  aagcctctgt ctcattctca aaaaataaaa ataataaaat ataaaaatgc ttatcatctg
```

FIGURE 4-H

```
26761  agctttcagt gagttttaat ctttttgctg atagggtatt ttgccttatt gttgatgact
26821  gctgactgat caggctgggg tgtctgtggc agtttcttaa aataagtcaa caatgaagtt
26881  tccctggaga atataaaatg ctgttgtata gcattttact cacagcagaa cttctttcaa
26941  aaccgcaatc aatcctctca aactctgtcg ctgctttacc aactaagttc atggaatatc
27001  ctaaatcctt ttttgtcatt tcaagggttc acaacatatt caccaggagt agattccaac
27061  tcaagaaatc agttccttcg ttcttccata agaagcaact cctcatctgt tcaaattttg
27121  tcatgaaatt gcagcaattc agtttcatct ccagactcca cttctaattc tagttccctg
27181  gctatttcca ccacatccag ttacttcctc cattgaagtc ttgaaccctt caaaatcacc
27241  cataagggtt ggaatcaact tcttccaagc tcctgttaat aYtgatattt tgacctcctc
27301  ccatgaaaca caaatagttt taatggcatc tagaatggtg aatcctttcc agaaggttta
27361  tttactttgt ccagtcccat cagaggaatc actacttgtg gcagctactg ccttacaaaa
27421  tttatttctt aaataataag acttaaaagt ccaaattatt ccttgatcca tgcatgagct
27481  gcagaatgaa tgctctctta gtaggcatga aagcaacatt aatctcctca tctttgtcca
27541  ttagagctgc tgggtggcct ggtgcattgc catggatgtg ctgtcaccca ggctttgttg
27601  ttctactaat agagcaaagt agatttagca taattcttaa ggactctagg atttctggaa
27661  tggtaaatga gcactagctt caacttaaag tcaccagctg tattagccct taacaagaga
27721  gtcacagcct gaccttggaa gctttgaagt catgcattga cttctcctca gctatgaaYg
27781  tcctagatga catcttcttt caatagaagg ctattttgtc tacagggaaa atctgttgtt
27841  tagtaaagac accttcaatt atctcagcta gattttctgg ataacttgct gcaacatcag
27901  accttgctgc ttcaccttgc tcttttatgt tatggaaaca gcttctttcc ttaaacctca
27961  tgaaccagcc tctgctagct tcaaactttt cttctgctgc ttcctcacct ctcagcctcc
28021  acagaattaa agagagttag ggccttcctc tggattaggc tttggcttaa aggaatgctg
28081  tggcttgttt aatcctctgt ccagaccact gaaacctcta tatcagcagt aaggcagttt
28141  ccctttctta tcattcatgt gttcactgta gtagcacttt tcttttcctt caagaacttt
28201  tcctttgcat tcacaaattg gctgtttggt gcccgaggcc tagcttctgg actatctcgg
28261  tttctgacat gccttcctca ctaagcttaa tcatttctag cttttgattt aaagtgaaaa
28321  atgtgtgact ctttctttta cttgaacact tagaagccat cgtagggtta ttattggcct
28381  aatttcaata ttgtgtctca ggaaataggg aagcctgaag agagggagaa agatggggaa
28441  ctggccagtc agcagagcag tcagagcaca tgaattttca ctttcttaag tagctgtggt
28501  ttgcggtgcc ccaaaacaat tacaatagta atgtcaaaga tcacctgatc acaaatcacc
28561  ataaaagaca tagtaatgaa aaagtttgat atactgtgaa aattaccaaa atgtgacaca
28621  gagacacaag gtgagcacat gctattgaaa aaaaaaatgg tgccaataag gcttactcaa
28681  caaagggttg ccacaaacct tcaatttgtg aaaaatgcag tatctccaaa atgcagtgac
28741  acaaagcaca ataaaacaag gtatgcctgt ataaatataa atatagaaat agatatagat
28801  gcctgtattg tatatggtgt gtacatatgt gtatgtgtgt gtgtatatat atatctattt
28861  cctatctctg tacactgaga aagactagaa gcaatggtat cccaaacaag gatcacgtca
28921  agtgcccaaa tcttggtttc taaatgccat cttccactaa aaagaaccag gtttcctgga
28981  gcagtaattg atcccagagc tggggcagga aaaagactgg aacatcttat gccaaaaaac
29041  aaagacagta ttcacagaat catgacaaaa gcacacagag accagctgaa agaaaatcSa
29101  gtgaccaaat ctatcacaat tcaagtatca taataaatgg tcaagatttt acaaccccat
29161  ggcataaaat agtagtagtc catactgata taaataaaaa gaatcatgaa taaaataatt
29221  ggcatgcggg agaaaaactg gcaactaatt aatgcggaag gaattaaaga aacagaaaat
29281  agaaaataat gtggtaccat cattagtggc tgataattca agtgggagtc ttgaatgtat
29341  gttaaggttg gtgaatggag gtttgacaag aaacaggcta ttaagagtat cagaatatca
29401  ctccacaata atatttatca attataagag aaaagcagtg tctttacagt agagaaacct
29461  ggcagacaat aacatgaaca ggtaatcagg attgatatca tgtcagcagg gcatggtggc
29521  tcacacctgt aatcccagcg ctttgggaga ctgaggtgtg cagatcactt gaggtccaga
29581  gtttgagacc agcctggcca acaaggtgaa agcctgtctc tactaaaaat acaaaaatta
29641  gcccagcatg gtagcatgtg cctgtaatca cagctactca ggaggctgag gcatgagaat
29701  tgcttgaacc tgggaggtgg aggttacaat gagctgagat cacacaactg cactccagcc
29761  tgggtgagag agcaagactc catctcaaac aaacaaacaa acaaaaaaca aaagcatcat
29821  gtagatgggt gaacgttgtg tgcctcctgt gtgatgcaca aattacatcc cctaactcct
29881  aaccctatag catcatttct gtagtattcc tgtcaaatat gaatggcaca catttgatca
29941  tgagttaagt taggaaaaaa actaagaaac tgttccatgt tgaaggaaac taagggacat
30001  ggcagctaaa tgcagtcgac gatccaggac tggatcttga accaggagga agaagaaact
30061  ctgagatcca ggactggatc ctgagctaga agggaaaagg aactgctagt gaaatctgaa
30121  tggagtctgt gaatgaacct gtgttgtatt aatgctgatt tccttatttg tatggttcga
30181  gggtagttat atccttgtct tgctgaaata cattatgggg catttaaggg aaataatata
30241  tcatgtctac aattccctct caaatagttc agaaaaaaga tgtatgataa tagacgtgta
30301  tttattagag gaggcaggga gggagatgga aaaatgttgt aaaatgtgga caattgtgaa
30361  atcttgatga aggggatatg ggatctctgt ataagactga aattatttca aaataaatta
30421  ttttttaagt gactaacaca agggacaact cctaaatttc aaagattaaa ggatcctagt
30481  aagtatggag atggaagaaa acaggttatc tacaaaggaa agaagaaata aggctgcctt
30541  caaatttcta ttcataagca ttaactctca tgaagcaaaa tagcaatacg gtgatagttt
```

FIGURE 4-I

30601 aaagaaaaaa agatcatgac tcaagattat gttctctatt tagctggtgg ttcattgtgt
30661 ttaaaaagaa aaaaaaaatc ctgagtcctt gctagaatat gtgagaatgt ctaccaactt
30721 tacaaagaag taacaagtat atctgttgac aataagtgta tatatattaa acaatttatt
30781 aaaggaaaaa gtactgacac agcgagtcaa aaacaattca atcatatgct gtttcttagc
30841 cacttttaat tctctaataa aagttaaata aaaagataaa gaaacaacaa ccaaccaaaa
30901 gaaagcagaa gtcacaatac taatattaga caaaacaaac tacaaagtaa aaagcattta
30961 acacaacgaa gcagataaat ttgaccaaag caggtacact ccatatgaag gcttatgtga
31021 tcagtaacag cactgaaatg tacataacaa agactacaga aaatacaagg gcaaagacag
31081 ttgcagtggg aacactttac ttcaactcta gcagtcctta atagatctca aataaaaaga
31141 atacataggt tgagaacatc aagcttagta tgtccaaatc taacRtcctg ttcttttcct
31201 caaaatctgc ttcttcccca gtagtcccca gatcaggtag taatattgcc attcttccag
31261 gagctcgggc caaaaacctt caagttctct ctaacgcctc tctttttctt atacaccacg
31321 tatcgtccat cagaaaatcc tgttgacttt accctcaaaa atacctaaaa atcctcctcc
31381 tgctttccac tttctactgc taccacattg gtctaagcca acacctctca ctaggattat
31441 ttcaacatct ctccagctag tctccttgct tccaaccttc ctccatccct tcaaccccat
31501 cagtcaactc ctaacacagg agctgcagtg attctgttaa tacaggtcaa gtcatgtctc
31561 tcttctgctc aaaaccctct aatggcttcc catcgctctt agtaaaattt tgcctttaac
31621 atcctatgtg atctttcatt ctgttgcctc tgtgacatca tttttcctac tactcgttcg
31681 ctctcttcct ctgctctaac cctctggcct tcttggtgtt cctcgatcaa gacagactct
31741 tgaccacata agaacttttg cccttcatct tccctctacc tggtacacta tttacccaga
31801 cagctgcacg gtttgctgtc tcatctcttc aaatctttgc tcaaatgatc acttccttag
31861 tgaggacttc tctgcccatc tacctaaaaa tgtaaacttc atttcccacc ctaaaagcca
31921 tgctttattt ctctactctc aacacccaac aaatcataaa tttgttatct ttttatattg
31981 tctgccatcc tgccccttcc agaaattaag ctggagggca gaaatttttg tcagttttgt
32041 tcactgacac acccctaccc ttaaaataat aagtaaatat tccataattt ttcatattta
32101 agaattccat aaatattaaa agtatattca aaaaaatata tccacttatg ttatcaacaa
32161 aataactgga tatatatctt taaatccata tctgtataca aattattccc tttttcaagc
32221 accaatggaa aaattataaa aactaagcat acactagaca ctagaacaca aaaaaaactt
32281 actgaattca aacaataaat aaaatagact aagttttcta gaacaataaa atgaaaccag
32341 aaaatactaa aagaacaaaa taaaacacaa cttcttagaa taacagcaaa aattggctat
32401 ttaacataca cctgatatta aactaagcaa agaatatgta ttcatcttta tatcagctta
32461 tgggattgga gctgatatca ttttacagat aaggaaactg aagtttaaga acactgaaat
32521 ttcttactca agataatata gtgatgcttc ccgatcagtg gtgaatgggg caaaaaaaaa
32581 aaaaaaaaag aacgtagcta gtaagtagta gagctagaac caatcccagg tccatgaccc
32641 tggagaccaa gctcttaact aatgtactct taaataactg ctgggtcaaa gaaaaatcac
32701 tacaatttag agtatactac aataacacta catagcaaaa tctatcaaat taagccaaaa
32761 ctatactgag ggaaattgaa tgctgcttct ttaccaaaga aaaagcagaa aaacacagaa
32821 agcattcaag ttaaaagaga aaatacaata aatctaatga cgggggaaaa attaaaaagt
32881 taaaagaata gtgattatta gaaaaaaaag tataacagat tacaatcaag cacgcatttt
32941 taggagcatg aaaaaattat aaaattctct attaagaggc atctttaaac ttcagtatga
33001 aaattttcaa gtactgaagg ccgggagtgg tggctcacac ctgtaatccc agcactttga
33061 gaggctgagg caggtggatc acaaggtcag gagtttaaga ccagcctggc caagatggtg
33121 aaaccccatc tctactaaaa agacaaaaaa attagccagg cgtggtggca ggtgcctgta
33181 atcccagcta ctcaggagtc taaggcagga gaattgcctg aacccaggag gtggaggatg
33241 cagtgagctg agattgcacc actgcactcc agcctgggag acagagcgag actccgtctc
33301 aaacaaaaaa aaaaaggaaa ttttcaagta ttgaaaatag agaaaatagt ataatgaacc
33361 caatacaacc atcaaccagc ttcagtaatc atcaacagat cactaatctg cttcatgtat
33421 acccttatta gcttaacaat tttaaatact ggagtagaca ccttaagtgc caatatttca
33481 ttatgttcta cttaaaaaac tgtcactcat tttctttctg tctcacttag gtacagggtc
33541 ttattaatct ctaactaagg gtcttaagag gtaggtactc aatatacact cgccatcagt
33601 gacaatgtaa acaaaacata gcactcctta ctccatattc ataatgccat aaaaatcaag
33661 acagaaaaaa aacaatagtt tacctgaact tccatgaact tcctcatcat ccacatcatt
33721 ctttccagac attaaataat tattattgag cctggaaggt ccacttaaga aaaaaaggaa
33781 accacaaaat taatgctttt ataatacata tatctaatgt atttgcctta atgatcattt
33841 ttctcctaaa actgccataa gtctttctta taaaaccaat agtgatgtta caagtcatca
33901 aaaatatata gtcatccctc agtatttgtg ggggattgat tccaggatcc actgtggata
33961 ccaaaatcca aggatgctca agtcccttat ataYaatggt gcagtatttg catgtgactt
34021 tcacatatcc tcctgtatac tttaaatcat atctaggtta ctgataatac ctaatacaat
34081 gtaaatgcta tgtaaatagt tgttatactg tatttttatt tgtattattt taaatgtcat
34141 attgttactt tttattgttt tatattttat tttctgaata ttttcaatcc atgggggttg
34201 aatcagtgga tgcagaaccc acagatatgg agggctgact gtacttgctg cattaaacaa
34261 agactacaat ttcatcccat aactgacagt tgagcatcat aagtactgaa tttatgactt
34321 acagatcaga gtaacagact tagtaaattc aaatgataac atcctgtggt gaaaatcaaa
34381 ttataaaaaa agaaagaaat catcatattt taattaaaaa aaaatatagc caaaaaccaa

FIGURE 4-J

```
34441  tggaatttaa aatgtggtcc agaaattgtc agagatgatt cagcctctcc tagtcctaaa
34501  tctctttcta cctattatga acccaggata agcctagtca taacctcaca aatagtaaag
34561  aaatcttagc catcccttaa aagtctgaga atttgctaaa atttaccctt tattcatttt
34621  ttaaaaggtt atatgacatc agatataatt aagagtaggt aattcaattt gatcactaaa
34681  tgtggtcatt ttaacattat atgtccaaat tctttaatac accacctttc aataggtgga
34741  gcctaattcc acctctccct gaatgtagtg tggactcaat aagttgtttc taatgaacag
34801  aataaagtag aaatgacggt gtacaatgtc aaagactaag tcattaaaag acactgtgat
34861  atccatcata gtcacacctt ctcttggagc atttgctttg tgggaagcca gttgccatat
34921  tcagcagctg agagggccac aYggttaaga aactaaggtc tactggaaac agccagaacg
34981  gaactggggt ctccagccaa cagcgatgtg agacatttta taggtggatt ctccagttcc
35041  agacaagact ttaaagggtt aaaaccagaa agatgactgc tacctcctga gaatatccga
35101  gccaaaacta gctagctaag cctcccagat tgctgaScca tagaaactat gaaacaataa
35161  gtgtttttaa gcttttaaca aaattaaaat aaagttcttg ggtaatctgt tatacagcaa
35221  tgggtaacta atatatacaa ctattatcct tacaatacac gcataccttt cctctaaaac
35281  tgaagtctta tgctctccag aatccaagtt tgactctgtt ggtatattac atgtatatcc
35341  actggtctga ggctttaggg gaacattctg tgcagtcata atgttatctt cattctttga
35401  gcttcttgta gatctataaa aatgataaat gtatgagtgt ttatacttct tcaaaataaa
35461  attttcttcc agaaacgata cacaggtggg tacaaagaca atgtaatatt ggctataacg
35521  aatatcagat caccctgttc aggttctgag acttgcaact agctaaggaa ggaaactggt
35581  ttctaggacc taccaagagt cccaatttga aaaattctga tgagagaaat ggagaagaca
35641  cctgccaagg aaggtggtga aagggacata aggagacctt atctataagc agcttgatat
35701  atattacagt taatctatca tggtttactg gagaatccat aaacaactaa aattggcttc
35761  tttgatatcc ctagtcagcc tactgagggg caatagggtc ctcagaggga ctgagctcat
35821  ctctttggta gttagtttac tccaaatata tgcaaacaag atacactctt cctagtgact
35881  gtataattct tcaggttcta gtatcctcct gggtcctacc attaaagtca aaagaaataa
35941  aagtgattac aagagaatac taggaacaac tgtatgccaa caaattaaat aacctagatg
36001  aaatggacaa actcctagaa agacacaaat tatcaaaatt gattcaagag gaaacagtct
36061  caaaatacca tgaatcagta atcaaaaaca aactacccat atagaaaagc tcaggcacag
36121  gtggcttcac tgcttttacc aaacattcaa agagaattac gacaaattat tcacaaaccg
36181  tttcaaaaaa tagagagata atacttccca actcattgta taaagccagt attaccctga
36241  taccaaaact aaacaaagac aacagaagca aagcaagcta ccgatcagaa cccctcaaca
36301  aaatactagc aaattccagc aacacaataa aataaaagag agagaaaagc agagaatggc
36361  ctgaaaaagt gaaacagaat gggtatggga tttctttatg cactataatg atatacgcta
36421  atatatttta ggacacactg tcatttattt cagtatatcc tcagcacctt actgattcca
36481  acacaaagta tgtattacct caataggttg ttggaactaa taagaattca ttaaaagYat
36541  ctagcaaata tttggcacat agtatgaact caaaaaatgt taactattat catcctaaga
36601  agaaaagtat taagctatat tcctaactat tgaaatcaca ttaatctcac ttaaaaggaa
36661  acaatttgaa ctatacactt aaaaaagatt aagacagtaa attttatgtt atgtgttttt
36721  tgtctgtttg ttttttgaat ctctactttg ggagcccaag gcaggagaat cacttcacac
36781  caggcgttca agacaagcct aggcaacaaa gcaagacctc atttctccaa caaattaaaa
36841  aattagccag gtgtggtggt ccacacctgt agtcctagct actcaggaag ctaaggcaga
36901  aggattatct gacccagaag tttgaggtta cagtgggcta tgatcacgcc actgcactcc
36961  agcctgggaa acagaggaag actctgtctc ttaaaaaaaa aaaaaaaaaa aaaaagttta
37021  actcttggat gacaataata tcttattaaa tttaacagtc ataattccag gtgctatgta
37081  ctattcccta ttttaatatg aggaaacaag tgcagagtac tactttgcct gtgtgtcaYc
37141  cagctaatgg tagaggctgg atttaatacc aggtatgtca gagtgcagag ctcttcatgc
37201  taattactgc ctgtccataa atgactaaat agcacaaatt attaaatgct gaacaacaaa
37261  agtaactata ttgtacacat aacagacatt acctagaaca atcaagattt ttcttcttag
37321  ccaagtctgc ctcatcactt tccaatggct cactcagYga acatctggaa atttcatcat
37381  gaccaacgat acctccagaa ccttcagcat ttaaaaactt ctgtactctt tggttgcctt
37441  tagttgctgt cttctgcctt ttaagtggaa tagtttttt cctaatattc ttagatgact
37501  gatttgtttt cttagtagat tttcggctac tattgtgatg cattggtaat tcatttctta
37561  ctgaagaatt gctataaaca ggacctgaag gattcaatga taacctaaag ttagtaagtt
37621  tgtccagatg atttggctca ctgaaatacc aacagcattt atatttgcat ataaacatat
37681  acacatacac acacacacac acacacacac acacacacac acacacacac acacatctca
37741  ggctaaattt atttccaaat taaattcttt aaacaatgta gctcttcagt ttatagatat
37801  atgaaacagg cccttcaaat ctgtcacagg cttcaagaaa cactctctgg ctctgcccac
37861  tgcccttctg gccatcctta atatctgata ttccccaagg tgacagcttc agcttttcac
37921  tccatatact ttcattgacg gacttatcta cccagagctt taattactat ctatatgctg
37981  acacctataa tatccatcta ctgctctccc ctatatctgc ctctcttgct ttatttccta
38041  tcttaatgaa aaatacaact acctattcta tccagatatc taaatcagaa atttgagtca
38101  tgcaagacta ctactcctat ccacaatcac gccctaaatc ccatcaattc tactttatta
38161  atgcctttca aatctttcaa gctctccctc ctgtcattag tagaagtcct cattatttca
38221  tatctggatt attaaaacag tctctttact aatctccccc tccctagcct tgtttttctt
```

FIGURE 4-K

```
38281  caatccattc cctacactgc tgcaagaatc acctttggaa aacaaaaact ggatcattac
38341  aactctgctt aattaaagta ttttagttat tatctatagc ttcaggatga agtccaaata
38401  cttcaacatc aattcatatc tacatctaga tttgcttatt tcataaacca caattcccaa
38461  aacccccctc caaacaaaca aaccaaaaaa agccatcagt aatgaacaat ctacagtttc
38521  tctaaagcac tggttgtcca aaggcagtct ctagaccagc aagatcagca ttatctggga
38581  atttgaaatg taaattatca tgcctcacct cagaccccag tgaatcaaag actttgcaga
38641  gcagggccca gtaacctgtg ccttaatagg ccccccagac aattctaaag cacactaaat
38701  taaagaaaga aaagatcaag ctctctgctc cagatctctg cacaagctac ttctacttaa
38761  aacattccat aacttcctcc ttactccact caatcactta gctaactcct agttcatcct
38821  tcaggaaRct ttccctaatt tccaagtctt aggtagcttg ctatcacaca tacgactgtt
38881  tacttatctc tacttctctc ctccagcaga ctaacaactc cttgacctca agtaattgat
38941  tggtatttta tctatgtata ttcccaatgc ctaatgacag gggcaggact atggtgaagc
39001  aagagtagcc tcagctggaa aattcaagaa ggtactcatt ttcaggtctg tgcaagtaca
39061  gagcccgaac tgatacaatc ctgtgactga ctgccttaaa tttttcaccc tagtgtctca
39121  ttgtcttacc ctagtcctgg tcctacttta cctcattcct gccccatggc aagtactcca
39181  tatacatttc ttcatatata tggaagggca acagtgtcac acgtacgaca ggcaggctgg
39241  gggcaggtca agttgttctt cttaacttca gacttcagtt tcttcttcca tcagatgaga
39301  gtttagacta tatgatatat aaagattcct ccacttttтc gactcccagc ttgatcaaaa
39361  atccaacatc tatttcagga ctcttaaata agtgtttttt gatagtttca gagatctagc
39421  acatttatcc aagttctgta agaatttatt tacccatctg tttatgtaYt catctgccac
39481  atagtactta gcatttcagg caacctgggg cagaaaagtg agttccccaa agcagtaaca
39541  gtattattta acacttcaat actaacatat gccacctact atgctaattg catttaaacc
39601  ttataaaaac tcaatgaggt aatcatgtta cagaatgaga aactaaacaa agcaatctac
39661  tcaaagtgga agagtgatac tttgaactca gatctcactc caaaatccaa aggctgcaac
39721  aaaataaata cttgctgcta tccatcaaga atctcaatga tttaataatt catagcaata
39781  aggcctcaca tatacaatgc catagttgta ctaatcattc acttattaat ctaaatgccc
39841  caaagtcaat agtatttgtg tatgtgaatg ctaccttatg atccaccaat cttttcagga
39901  tcactttctt taaaaaggga caattctccc aaaacaataa atgtttcata ttgtgcagaa
39961  aaaagttagc aaagcaggtc tgagtatact acccttaaaa gggcctgctt tcaaagttag
40021  ccctgtaaac ttaaatttca ggattgttcc cgtcatttcc taattggcaa acgtggtgtt
40081  gtgggcctaa actgcttatg caaaaaatat ggtttaggca aatattgcct tccttcaggg
40141  agtctggaat tgtagtactt gctagataga gagtacctat gtgaccaccc ccaggaaaga
40201  tcttgggcac taagtctcta gtgagcttcc ctRatagaca ttttgcctat ttgtcacaac
40261  tcattgctag aggaattaag cagctcctgt gggaatccac tgggaaagga ctctttaagc
40321  ttctgctgat ttcctccaga cttaacccat gcacttttтt tctactgaat ttgcactgca
40381  tcctttcact gtaatgaatc acagccatga gtacaactat atgcccagtc ctcctagcaa
40441  attaccaaat ctgagggtag tcttgggaac cWccaacaca agtaatttac ttttttatgt
40501  gtaataaact ataatctttt agattataaa ctccttggtg gcagaacctc atattatcaa
40561  tctttgtatc cctcacagcc aggacattaa aatagctgga ttatttatat tgcctttccc
40621  aactagaatg caagctccat gaaagcaggg atttctggtt tggggttttg ttttcccaca
40681  gctattatcc ctattcctaa aatactgtat atattaacgc ttcaaggaat atctttacca
40741  ttgatttaat attccttttg atgacccaat ttaacttggt cctcatgata ttctctatat
40801  agaaatataa aatcaaacat aatgcccctt acagaatgat tttgtctatg aacaaacaca
40861  atttaaataa aaatacttac tctcctctga ttttaccacc caccaatcag atggacgcct
40921  ggaaattctt cgacttttcg tgacagttga agtcacttct tcaggtacaa gtttgttctt
40981  ggactcactg gaaaagcgct tctttttaga ttcttcctta tcttttctag tgctactttt
41041  cttgcttcct aaaataaaga aaaccataaa accaattcac aatctactaa aagagttttc
41101  agaaaaccac aagattataa aaaaaccgtc aattacagga acttcttaaa tttatcctgt
41161  ggccttttca tctccatttt taaaagtcac atgttctctt tctgtcaatt atatgtgctg
41221  ctgaagtaag cactaacttc ctattaaatg caaattttca cactgcttga taactcttat
41281  atactcctca gccaaaaaaa taagtatctt ctgaaactga aaaaaaaata cagttccttc
41341  ctataacgtc ctaagtaaaa gtaggcaatt actgctttcc tgttttgaaa taatgatttt
41401  tttcaactta agaagaggag atttaaagtg tgattaaatt ataaactctg aaatagaggc
41461  attatcttgg atcatccagg ttgacccaat ctaattatgt ttcttttaaa agctgtggtc
41521  agagagatga gatgatggaa gaagaggcaa gataaattca caacatgaga aggctggtat
41581  tgctagtttt gcagatggtg gaggaggacc atgagtcaag gaatataggt agtttcaaga
41641  gggtggaaaa ggcaaagaaa cagacttcct cccagagcYa cYagaaagaa agaaaacccc
41701  accaatacct tgattttatc ccagtgagac atctgaccca cagaactgta agataatcaa
41761  tgtttgttgt tttaagccaa taagtgtgta gtaacttgtc atatgacagc aatagaaaat
41821  taatacaact gatatggtaa ttaatgagca ttatattcac taaccaatct agatacagaa
41881  tatatgtaaa aattagtgtg tagaatatag caggtactca gaaaatgttt aatatttaca
41941  attatagagc gcatctttca agtatacttc cttcatatgt cttttcctaa gaaattatgg
42001  ataaatcaaa attttgtatc cttcctttta atcctatgca tgtatgtata ttatctatta
42061  ttctacagaa cccaccagaa tgatcttgtt aacagtagaa ataatgggtt agtaattttt
```

FIGURE 4-L

42121 gtgattacag tgcctaatgc agtgtatggc attcaacaac tgttagacac atgagtgaat
42181 gatggaagag aatggacttt tatctctata aaaaatagtc ccacggtacc ctgggatttt
42241 ttttaacatt tttatttttct aaataattta tcattctgtc ttattcatct tccgatacta
42301 acttgtttac tagccttttc ttttcacttt gccctctata tatactgcca tctcacaaga
42361 taaacaaaaa ctatatgagt tagaattaaa ttatatataa aataaaaagg ttctagaaaa
42421 gaacctatag gcttttcagg ctctttgttc atttaggttt actttgccta cttgccaaag
42481 ctaagttagg atattaaaaa acataaaatg ggtagaataa tgctcatggt tatatttaat
42541 ataaacactt acccacaggt ggcatctgtt ttttggaaac acaatcattt cccatctctt
42601 catgctcttc catatttctg tctgaatttc tttgaaattc gtcttgggta atatgtgatg
42661 tatgtatgtt ttcatcttta gactgtccca catcaagctg ttcttcagct ggtttagcca
42721 tgaattttct tctctgtttt tgttttatag tccttttgct tctagatggt ttttctgcat
42781 tcttggaata catttcatat tttgtagatc tataattatt tactgtttca Yctatcaaag
42841 cataacttgt atccagtact gttttatcag agggctgaga tgtctctact gggtgaggtt
42901 tatgggaatg tttgtcattt gccaaagtct taggtaatat attatgatgc ttttctcttg
42961 actttctacc ttgaaRgagt gcagtgctct cagccgggga tattgtgcgt tgtttcagag
43021 accctgcctt tcttggtatt gtaatccaag atctactggc aaaactttga tccgactcat
43081 caattataaa ttcatcctct atcaacttcg tatcatcggg aggacacgaa tgaggtggag
43141 cagttgccgc atgcctaaca atgggactga aacagggtga tacttttaac agttcaaaaa
43201 aataaagctt ttatatttgt ttagtaaaga aaataatcta ggaaataaaa tctaaatttc
43261 ttttaaactg tttatatatc tcctcaattt tcccttaatc ttacaaatct ttccctcctt
43321 tcacatttct atgtctgttc accttattt ataatagtaa tgacatcaca aaaatttaga
43381 gctagaaatc tagacatcaa caaatttgat cccacccacc cacccattat atatactaaa
43441 actgaagccc agagagatat aattataact cagatacatt gactcacagt ccaaaatgaa
43501 tgccactatt atatataatc ttatattaaa ttatatattt gtatatataa cttttggggt
43561 ctgaatttct tcacattcaa aataagtcaa caagattatt tctgagaaca ctccaagtct
43621 aatatttgaa tggactacta ctaaattata aaaagttgag gaggccggat gtggtggctc
43681 acgcctgtaa tcccagcact ttgggaggct gaggcaggtg gatcacttga ggtcaggagt
43741 tcaagaccag cctggccaac atggtgaaac cccgtctcta ctaaaaatac aaaaattagc
43801 caggcatgat ggcgggtgcc tgtaatccca gctacttggg aggctgaggc ggaagaatca
43861 cttgaaccca ggaggcagag gttgcagtga gccgagattg caccactgca ctccagcctg
43921 ggtgactgag caaaactcca tctcagacag aaaaaaaaaa agagagagag agagagattg
43981 gggacaattt cttcctatct ttttttacat cccacagttc ttcatacata gcaacaccaa
44041 atacacagac ggttgaatga atgtaaaaat tactaatatt actataggac tgtaaattgt
44101 agacctacag aactataaaa taataaaYgt tgttttaagc cgctaagtgt gtagcaactt
44161 atgacagcat gtgtgtaaaa atgcattcca agttYgggta tttcacatat attttcatat
44221 tataatcatg gtatagaaat aaatggagtg tatctagaaa ggtagcataa gcacaaaaga
44281 aatctagtgc cacagaaaca ggaaaaataa aaggcaagtc ttatcacaaa gcatacaatg
44341 ctacagaatg ttattaaata atgaagtcaa acaatgtttc ccaactctag acctttctcc
44401 atagtcctag cccaaacttt atgtatttta ctaaaaggca gaaagcctat atcatgaata
44461 ttattgttat ttcccatgcc cagtgttgta aaaggctaaa gagtaagaca gacaatttca
44521 actacaatct tatggcttat aatataacat caacattaat tcttaaattt tcataaaagc
44581 atgtacatgt taagaacata caaagagatt catgcattta atgaggtgtt gaataaatta
44641 agactagtgc attgagatta cacagtacct cctccaatta gtcaggaaac tttcctagag
44701 aagggaggaa ttattccaag gacagtaatc tgatagactg actaatttta catttccctg
44761 attgttttaa tagctttaag cacaccacat attgcagtga gatacacaaa ataaaaaaca
44821 aggtaaacat ccaaactcca aagctggcat acgcactcgt ttccactttt tcactaggat
44881 aagaaggctt agttagctat gtttagccat aacagaaaca aagacagaat gcaaaaatta
44941 gattatgttt ttctctttta attagaacat acacaaggaa tcagaaatga aaatgccttc
45001 tctattcctc taatcgcaaa ctcctgacaa atcttataat gtaaaatagc aaatactcta
45061 aaaaatatac tctctaatgg tagagtttca caaacttata taattagaaa actctataaa
45121 aactaataaa gaacaaagtc aatacaaata ctacaatgaa tggatgttaa acagaaatga
45181 tatccttagg aaacaatctt gacactatga aagtttacaa tagggaaaga tggcttctaa
45241 ttagctacaa aaactatgaa acgctttgta gcatttaggc tagatcttgg aggatttcta
45301 agatttggtc acaaggaaat gcaagaaaga caatcaaaag gaagagcata agtggaagca
45361 atgatgctag aaagcactgg gaataatgaa tagttatatt tgactacaac aattaaacat
45421 aagatgaact ttactcacta gagtatcaaa aagaacatta aaaaataatt tctcagacat
45481 tatataatta gcccacaagt ataacgcatt tgtttcagat ggtggtgaca atatccatat
45541 gaggtttaaa atatacacat aaaaataatt aatcaactac aggaagcaga gactgaacaa
45601 aaggggaatt tcattcattc atttaacatg tttttataaa aacttccttt ttatttttga
45661 gactgagtct cactctgtca cctaggctag agtgcagtgg caccgtctcg actcactgca
45721 acctccacct cccaggttca agtgtttctc ttacctcagc ctcccgaata gctgggatta
45781 caggcatgtg ccaccaagcc cgactaattt ttgtatttttt agtagagatg gggtttcacc
45841 atgttggtca ggctggtctt gaacctctga tctcaagtga tccacccgcc tcggcctatc
45901 aaagtgctag cattacagac atgagccacc gcgcccagca ataaaaactt cctagtatat

FIGURE 4-M

```
45961  tcactagcac tctgatacct atgatagtga cataagggca gcctaacata acgatttcta
46021  aatactgcat tgtcggccgg gtgcggtggc tcaagcttgt aatcccagca ctttgggagg
46081  ccgaggcaga tggatcacct gaggtcagga gttcgagacc agcctgacca acatggggat
46141  atcccatctc tactaaaaat acaaatatta gacaggcgtg gtggcaggtg cctgtgatcc
46201  cagctactcg ggaggctgag caggagaatc tcttgaaccc aggagggaga ggttgcaatg
46261  agccgagatt gcgtcactgc actccagcct ggaagacaga gcgagaccct gtctcaaaat
46321  aaatacatac atacatacat aaatactgca ttgacttcag tctttgatat cacacagaca
46381  aaccagaagg caatgaaaac tgaataaact tgttttcttt caagtctaat ttagtatcta
46441  ttaaaaaatg gctttgggag tgactcactg actggtaaaa aaccatctct ccttacaggt
46501  cagaaaaaat ttcaattaaa atttttgtta attaaattac ttatttctaa gttaaatagc
46561  attaggaatc cttatatatg gttcatagaa tataaaagaa aaatgtctca aagggcagtt
46621  tcttaataac tcacctggat tcacttttтc gttttactgt ttctaaaaac aatgttgaaa
46681  aacttttttt agcttgtcgt tgaatttcat attctgaatc tcgtattcta ttctgagatg
46741  atcctgatgg ttttctttct tgtccttcYg atgttttatc ctcttcatct gatactttat
46801  tatctatttc tattttcttt aacataactt tatcatcaaa gtttaaccta aaaggttaaa
46861  aggagggtaa gctgaatgct cccgtaacgt ttcttgagtt ataagtcttc ctgaactcct
46921  ttacagagac aatacagacc aatttaaaat tctgtactag gcatttattt aagacataaa
46981  tgctgatttg tgcattccaa atacataaga taaactgctc aggaaaatta acttttgtat
47041  ttatcaatag aaaaacatga aagagaaaat aatggctaaa gtatacatca atgaaatgta
47101  tagctaggca atggttctca gggcaatctt agagcactgt gacagaccaa actttgaata
47161  tcaaagtctt agaatatact taactccttg gctccagtaa tcatacaata ctctaaatag
47221  aaattacact cattacagag atctccaaat gcaaaagacc ttgaatccac tttgtttcct
47281  aaaggccaaa ggtaaggtca agaaagagaa accagctgtc cacaaagcat agcatcctca
47341  ataatactta tgctctgcta ccSttgaagt gacctgatac acatctacac ttgcaattac
47401  aaaaaatatt gagaaaatac aaattgtttc tcttaggaag caatcaaata ttatcagaga
47461  gaaaacacat tattccagac tatagaaaag ttaacctcta aacttctctg aagtacactt
47521  taagcccttc cttaattatt acaaactgaa tatatgctaa atgtatactt tctcataaca
47581  tgttaatatt cctccttcct tagtgtaaaa taatataaca agattgaatt accatttttt
47641  aatacaaaaa gtaaaatttt taaagtgcgt taacattatt taaataaaat gttacctttt
47701  tttggtttta actgaaacct ctgtacttga aggcagcata tttactgaat tttcaaaagt
47761  gtaagtctct ctcttttggg cactagatgg aataacattt tgtgatacag atgtttttgc
47821  atccaaaaga acagaaggtg agccaacgga taagtaaaat tcttcatcag cttcactgtg
47881  atgatcattt atgtttctac ttgatatttt tttcgagtca ggtgtatttt tggaactaac
47941  atcagttgcc agaattttct gatgaacttc atgggcctga actagaagaa aattatataa
48001  agatatttgt caattaaaag aataataaga gaatcaagca ggctttttaa aaattctgca
48061  ataatttaat cagaaatagt tttcagaaaa ctcaaagtca ttaagattag aatcaaagat
48121  ctatgattaa aataagtaag tacttaagat tcttctatct ttgggctaaa aattggtgac
48181  ctagggcagc tagatgatat atttacacat acacacatgg aatatagata taacagatat
48241  agagagatac aacataagca tgtgtgtata tatacataca cagtatatat gtatatgaac
48301  atttatctat aatgagaaaa tttaaaagaa tgtatccata ttctcttaac cctcaccctt
48361  cagaacgaga tggcatttcc ctgcttctcc ctctaaatac aactaacttc cctggaaatt
48421  atccatcaat caacaacagg actctgaaag ctggaagaaa aaaaggtaga ctagctagga
48481  agctcaagac cagaaacaga caagccttag gttttctgtc tgtctcccat ataccctgg
48541  actggttgct agtataacct gtaaccacca ataggtgaaa gacagggtgg aggctgggtg
48601  gaaatgcatt ctctgaacaa ggtaaacca ccgaagacct aatgaggaaa ctcaagcccc
48661  attcagcact ctcggccatg gagcactcct tactacctgc attggcaaag caacaaagcc
48721  ctgggctgtg ctatgggact gtagcagtaa aggagctcat accaagccag tcccgcttct
48781  acaatcatag atggcaggcg gtatgaccca catgaagctg tagagtagtg aaacccctga
48841  ccYtcacagc ccaaggcaac attttgatcc actgatagta ttgacacccc aggccatccc
48901  agtccctact ttacaaccag ggcaccagca gacagtctaa tctgaggcaa gaacagcgta
48961  agagaggcca gcaacccccт gtcctccacc caatgatata aggaaaccca ggaccagcag
49021  cttctaacca gaaccccaca caagtgaaca tggacatcaa aaggcatctt ctacccaccc
49081  cagacagcac caacaggaac tgagctagaa tcctaacagc atcagaaaat gaagtagaac
49141  aaaataccac tattactctg aacacttaac tatcattgaa aataaaactc ataaaactag
49201  gaaagagggg ctgggcgcag tggttcatgc ctgcaatccc agcactttgg gaggccaaaa
49261  cgggtggatc actggaggcc aggagttcaa gaccagcctg gccaacatgg cgaaacccca
49321  tctctactaa aaatacaaaa attagccagg cgtagtggca catgcctgta atcccagcta
49381  ctcgggaggc tgaagcatga gaatcacttg aacccaggag gcggcggttg cactgagcta
49441  agatcacacc actgccctcc agcctgagcg acagagcaag actctgtctc aaaaaataaa
49501  caaataaata aataaataaa taaataaata aataacacag gaaagaacct atacattaaa
49561  ccaaaatagg gcaactatgt gctgtaacaa aagatttaag tagtctaaga atcccctaac
49621  ataagtaacc aaatctcagg aatacaatca aaacaactcg tcataccaaa accaaaacaa
49681  accaaaaaaa aaaaaaagta agaaataatg atcaatagac ataaatgaga tgaatcagat
49741  tttaaaatta tttaacaata aattttaaag caaacatcat aaaatgcttt aataatcaat
```

FIGURE 4-N

```
49801  tacaaatcct cttgtaacaa atgaaaaaga aactctcaaa gtagcatttt ttaaaaaaaa
49861  gtttttgga tattagccat tatcagtaag tagtgatttt taaatattat tttaaatttt
49921  cctaataaca gtgatgtaaa acaaatcaaa actacaatga gatactacat cacacccatt
49981  aggatgacta ttatttataa aaacaaaaga taacaactgt tgatgtagaa aaaaatggaa
50041  cctttgtgca tgcactgcta gtgtgaatgt agaatgacag tcactatgaa aagtgcacag
50101  tggatcctta aaaacttaaa actgaaagca ggaactcaaa gagatacttt cacactaata
50161  ttcatagcag cattactcac aatagccaaa aagcgaaaac aacccaagtg cccatcaaca
50221  gattaataga taaaaagtgg tatacacata caatgaaata ttacccagtc ttagaatgaa
50281  attctgacac atactacaac atggatgaac cttgaagaca ttatgcttag tgaaataagc
50341  cagtcacaga taaattttct gtctgaaatt catggacaga aaaaagtaga atggtggttg
50401  ccagggctaa tggaaaaaga gatactgttt aatgggtaca gtgtttcagt tgtgaaaaat
50461  gaaagaagtt ctatggattc atggtggtga tagcctcaca acaatgtgaa tgtacttaat
50521  gctactgaac tgtatactca gaaataaagt ggacaatttt atgtttatat tatacaatat
50581  tttaaaatac acttctttaa aaaagagcaa aggaagtgac gtgtgtatat aaacaaacta
50641  aaacaaaaag tcttgtaaag ttaaaaaata ttaaaatacc tgacaacaga aataactaac
50701  agtagtattt aagttgggaa ttcgttgttt cagttaccat gtctgcataa tataacttct
50761  taacttggct gcataaaaca actatttgtt atgttcatgg attcagtaag tcagcaattc
50821  agcagagcac agaagggacg gattcttcca tgaagtctga gccctggaat catctgacag
50881  cttgtttaca catatttgac acctggtctt gactgcaggc tgggagtctt ttctctccat
50941  aatggctact ttgggcttct tcacagcata ttgtctgggt tccaagggca actggagaga
51001  ataagaaaga caggaagaga gagagagaag gagagagaga tccagcatgc caagcagagc
51061  catatagcct tttatgacct aatcatggaa atcacatgta ttacatatgc cacattcaat
51121  gtatatgtag aaacaacctc aatgaagtgg atgggggggaa aggtgatgac ctaagaaaga
51181  tcagaaatga atcaaaggcc aggtgtggtg gctcatgcct gtaatcccag cactttggga
51241  ggccaaggtg ggtggatcac ctgaggtcag gggttccaga tgagcctggc caacatggtg
51301  aaaccccatc tctactaaaa atacaaaaat tatgtgggcg cagtggtgtg tgcctataat
51361  cccagctact tgggaggctg agacaggaga atcacttgaa ctcaggaggc ggaggttgca
51421  gtgcaccaag attgtgccat tgcactccag cctgggtgat aagggcaaaa ttccatctca
51481  aaaaaaaaaa aattaaaagg aatgaatcag ctgggcatgg tggctcacac ttataatccc
51541  agccctttgg aatgctgagg caggcggata acttgaggtc aggactttaa gactggcatg
51601  gctaacatga tgaagccccg tctctactaa aaacacaaaa aattagatga gcgtggtggc
51661  tgctgcctgt aatcccagct actcaggagg ctgaggcagg agaattgctt gaaccctgga
51721  gtcagaagtt gcagtgagcc gagatcacgc cactgcattt cagccaggga gacagaggga
51781  gactctctca aaaaaacaaa aaatgaatga cgtctgtaag actgaaggca aaagaaactg
51841  tacataagca ctattctcta gttgataaaa ctgttttcca caggatatga gtcaacaatt
51901  ctgacactct tttacatgta tacagaaatg taaaagataa gtaaatggat gttgaatggt
51961  gaacgcccag tttctcattc atgtaatggg agtttacaga taagcaagag gaggaggcta
52021  gaataatcca cgtggtaatg gattagagtc agagacatca atatgaactt atgattactt
52081  caacatagat acagatagtt acatttaaag atatttatag gtatgtgttg atacacaagt
52141  tattgtacac acacatacat ttacccagct ctatcagttg aaagggtcta gcagcaacaa
52201  ctctacagga gtgacagcac acctagcact cagatcttgg tttctaataa cattacccag
52261  taaaagaaat catgattcca tgaagaaatg gttgattctg ggatggggca ggagagcatc
52321  ttgtggctct aaaaggtatg aaagtgctca gaaacaacaa caaaaatggg ccatgtcaaa
52381  gggacactgg aaccaaagta agagcttcta ttagccaaat ctgcaacaat ctgaataaca
52441  aaataaacag taatgaatta taaagtagaa aataaaattt atccatcaag aaaagaacaa
52501  aataaaccca atcaagcaga aggatgggaa aaaataaaga gcagaaatca ataaaactga
52561  aaatagaaaa ataYagaaaa ctattgcagg aaaaaaaact aYaaaatgat aaacctctat
52621  ctaacaaaat tgacaactat aagaagacac caaccatcaa tatgtggaat aaaatagggg
52681  atgatgtcac aatagatttt gcagccacca aaaatagggt aatgcttctg tttgaaaact
52741  ggcacgagac aaggatgccc tctctcacca ctcctattca acacagtatt ggaagttctg
52801  gccatggcaa tcaggcaaga gaaagaaata aagtgtattc aaacaggaag agagaagaga
52861  ggaagtcaaa ctgtctctgt ttgcagatga cgtgattgta tacttagaaa accccatcgt
52921  ttcagtccca aaactcctta agctcataag taacctcagc aaagtctcag gatacaaaat
52981  taatggcaaa aatcacaagc attcctatac accaataaca gacaagcaga gagccaaaac
53041  atgaatgaac tcccattcac aactgctaca gagaaaataa aatacctagg aatacaactt
53101  acaagggacg tgagggacct cttgaaggag aactactaac cactgctcaa ggaaatgaga
53161  gaggacacaa agagaaaaaa attccatgct catggatatg aagaatcaat aYagtgaaaa
53221  tggccacact gtgcaaagta atttataaat tcaatgctac tcccatcatg ctactattga
53281  ctttcttcac agaactagaa aaaactactt taaatttcat atagaaccaa agagcccata
53341  tagccaagac aatcctaagc aaaaagaaca aagtggaagg catcacacta cctggcttca
53401  aactatacta taaggctaca gtaaccaaaa cagcttggta ctggtaccaa aacagatata
53461  tagaccaatg gaacagaaca gaggcctcag aaataacacc acacatctac aaccgtttgc
53521  tcttcgacaa acctgacaaa aacaagcaat gaggaaagga ttctctgttt aacaaatggt
53581  gctgggaaaa ctggctagcc atatgcagaa aactgaaact ggacccccttc cttacacctt
```

FIGURE 4-O

53641 atacaaaaat taactcaaga tgcattaaag acttaaattt aagacctaac atcataaaaa
53701 ccctagaaga aaacctaggc aataccattc aggacataag catgggcaaa gacttcatga
53761 ttaaaacacc aaaagcaatt Scaacaaaag ccaaaactga caaatgggat ctaattaaat
53821 gaaagagctt ctgctcagca aaagagtgaa caggcaacct acagaatggg agaatatttt
53881 tacaacctct ccatctgaca aaggtctaat atccagaatc tacaagaaac ttaaacaaat
53941 ttacaagaaa aacaaaaaca aacaatccca tcaaaaagtg ggcaaatgat atgaacagac
54001 acttctcaga agaagacatt tacacagcca acaaacataa ttaaaaaaaa gctcatcatc
54061 actggtcatt agaggaatgc aaatcaaaac cacaatgaga taccatctca caccagttag
54121 aatagtgatc attaaaaaga caagaaacaa cagatgctgg caaggatgtg gagaaacagg
54181 aagactttta cactgttggt aggagtgtaa attagttcaa ccattgtgga agacagtgtg
54241 gcaatgcctc aaggatctag aaccagaaat accattcaac ccagcaatcc catcactggg
54301 tatgtaccca aaggattata aatcattctg ctataaagac acatgcacat ctgtgtttac
54361 tgcagcacta tttacaatag caaagacttg gaaccaaccc aaatgcccat caatgataga
54421 ctggataaag aaaatgtgac agatatacac cacagaatac taagcagcca taaaaaagaa
54481 atgagttcat gtcctttgca gggacatgga tgaagctgga aaccatcaac ctcaacaaac
54541 taaaacagaa aaccaaacac tgcatgttct cactcataag tgggagtaga acaatgagaa
54601 catatggaca cagggaggga aacatcacac accagggcct gtggtggggg gagggaggaa
54661 agggaaggga gagcattagg acaaacacct aactcatgtg gggcttaaaW ccaagatgac
54721 gggttgatag gtgcagcaaa ccactatggc acaggtgtac ctatgtaaca aacctgcatg
54781 ttcagcacat gtatcccaga acttaaagta aaaaaaatat ttaaaaaaac tcaaaaaaaa
54841 aatagggtag tgctatgaac aactttatgc ttacaaactt gacagatgat tagaaaaaat
54901 ggaccgatac ctcagaaaac acaaactacc aaaattcagt cacaatgaaa ttgataatca
54961 aaatagtcct ataactatta aagaattaat attgctattt aaaagttctt gaaaaaaaaa
55021 tctcctggac cagatgattt caccggagac tgttacacca aatagttaaa ggattaacac
55081 caattttaca cagttccttc cagaaattag aaaagggaat aattcccaac tcatttaata
55141 aggcaaatat taccttgata acaaaaccat acaaacacag cattaaaaaa actataaacc
55201 aatctctcct tacttaaata caaaaatcct caaaatttta ccaagtcaca tccaacaatg
55261 tataaaaaat agcaatagtg cacaataatt gacatttata gataactaca cccaacaaca
55321 gaaaacattt tttttaaatg cccattcata aagacatgca atatcttggg tcataaaaca
55381 aatcttcata aatgtagaac tgaaattgca cagtgcatac tatgtcataa aggaatcaag
55441 atcagaaatt aataaggaaa agataatatg aaattaccaa acactggaaa attaacacac
55501 ttctaaataa ccaataggct aaactctgaa aagatattta aaaatacatc aaaataagtg
55561 aaatgaacac aaaacatgag atacaaaacc tgtggaatac agctacaaca gtgctgaggg
55621 ggaaatttat gccaaatgaa tacatttgaa aagggaaaaa gagctcaaat cagtaataca
55681 ccatgaccac atggaaattt tacgcaaggc tggttaaaca ttcaaaaatc aatatgtata
55741 agccaccata tcaacagttc agagaagaaa aatcatatga tcatatggat ttatgcaaaa
55801 aatgcatRac aaaatccaca ttacctatga ttaaaaattc tcagcatact ggcaggagaa
55861 aatttcccaa ccctataaaa aacatctatt aaaaacttat agctaacatc ttaatgtgaa
55921 agatcaaatg ttttctaaga gcaggaacaa cgctagaatt ttcaccccat ccttttttt
55981 ttttttttt ttttttttga gacagggtct gactctgtac ccaggctgga gtggggtggc
56041 actagcacag ttcattgcag cctcaacctt ttgggctcaa gtgactctcc cacctcagtc
56101 tcccaagtag gcaggactat aggcacatgc cactgcacaa gctaaatttt tgtgtttttt
56161 gtagagacaa gttctcacta cgttgcctag gctggtctca aacatctggc ctcaagagat
56221 tctccaatct cagactccca aagtgctgga actataggcc tacgccacag tgcttggacc
56281 catcactgtt atttaacata atactgaagt actagctaat aatagtaaga aaaggaaatt
56341 aaagacataa actaaattta ccccaaacta cccacaatta ttattttaag tggcacagta
56401 ctaatttcag ttaaaatcat agaacagaca tggataccta atatcatcat tattaaacaa
56461 cactgactca gagattctag caaaggcaag aagactgaaa taatcagatt aaacataaca
56521 aataaaacct ctctttagct gatgacatgg ttatatagct taaagaaaac taaaagattc
56581 aaatttttaa aggtctagaa ttaatttttt taaaaacctt ggtaagtgct gaattaaaat
56641 tttgatataa aaattagcct cataatttaa aaaaatgttc agaatagtac aaaaattata
56701 gtacagttag gattaacctt aacaagaaag tcaaagaccc aatactagga gggaggggag
56761 aatggggagc caggagacag gctttattaa aatatttttt atgataaaac atgaaattta
56821 caatattaac catttatgag tgtacaattt aatggcatta agtacattca caacactgtg
56881 caaccatcat catttctaga acttttttcat catttcaaac aaattctgta cacatcaaac
56941 attaaatcac cattaccccc taccacaatc ccctgcaact tctattctgc ttcctatctc
57001 taaatttacc tattataaat acctcatgta agtagaacta caaaattttt gtccttttt
57061 gtctggttta tttcagtttg catgttttaa tgtgttcaag attcttctat gctatgttat
57121 atagcagaac tttgttttta atgactgact aatattacat tctgtgtata aatcacatgc
57181 tgtttatcca ttcatctgtt gacggacact tggcttgttt atacctttg agtattgtga
57241 ataatactgc tacgaatatt agtatacaaa tgtctgcttc catccctgct ttcaattctt
57301 ttgggtatat acctaggagg gaagtattgg gtcatatggt agttctatgt ttaacttttt
57361 gagcaactat caaattgttt tctatggtgg ctgcaccatt ttacattccc accagtgatg
57421 catgagggtt ttaatttatc cgtatcttca acacttattt tccttttttt gttttaaata

FIGURE 4-P

57481 atagctaacc tagtgagtgt aaagtggtat ctccctaggg tcttgatttg cattctctaa
57541 tgactgaaga tttcaagcat cctttcatat atttattggc caattgcata ccttctttgg
57601 agaaaagtct atacacagaa tgtaaagtcc tttgctcatc ttttatttgg gttgtttggt
57661 ttttctgttg ttgagctgta gaagttcttt atatattctg attataaatc tctatcatcc
57721 ttttcaggta aattatttgc aaatattttc gcctagtctg gttgcccttt cacactcttg
57781 acagtgtcct tcagatttat ttatttattt atttattttt ggagacgggg tctcactctg
57841 tcgcccaggc tggagtgcag tggcgcaatc ttcggctcac tgcaacctcc agctcctagg
57901 ctcaagcgat tctcctgact cagcctcccg agtacctggg ataacaggta cttgccacca
57961 cgcctggcaa attttggtat ttttagWaga gatggagttt taccatgttg gccatgctgg
58021 tctcaaactc ctgacctcaa ctgatccacc tgcctcagcc tcccaaagtg ctgggattac
58081 aggcatgagc aaccactcct ggctccagat taacattttt ataagaaaga aacataactt
58141 attttgtaaa aaacctaaat aacataaaaa atgtgaattc tcacgaaatt actacataaa
58201 ttaaatgcaa ttctagttag aattccaaaa gtttatttgg aattggataa aactaaaagc
58261 tcatagaaag aaaaaaggct gcagaatggc agtggaatag tgaaggtatt tatatcagga
58321 catactatga agctactatt agtaacaaga gtattgtgtt gacagaggaa aaaacaaata
58381 tatcaaaatt caaaagtatg ttgaatgtta tgtgagattt aatatataat aaaagtagta
58441 cattcagaag agaacagatt tttttttaat ttaataaaca gatctgacac aactgaccaa
58501 ctggaagaaa acaaacttag aagattatgt catgttacat accgaaacac attccMgatg
58561 gttaaagact taattttatt ttttaaaaaa aggttttcag atgaaaaatc taagttggga
58621 taagacagaa cttttaacta aaacaaactt aaagctataa aagtcacaca tttaacaatt
58681 aaaatccaac agccaataaa cataaaggtg ctcaaactct ggcaatcatg gaaatggaca
58741 ctaaagtaaa aaggagattg cttaatatca atgaatgggc aagaacgtag aatagtaaat
58801 agactatcag gtaaaaggta ctcttacact attggtgtaa aaggaaatta taatagcctt
58861 ttctgaaagc aggtacaggc acagacacat acaaacacac atgctttttt ttttttcatt
58921 tatgggcacc aattaaggta cagaactatt taaaggctac ctcagcagaa tgataggagt
58981 aatgactaaa ttcttatttt aaacctttat taactcaatg gagtaacact cttttaatct
59041 gaaaaataaa aagaaagaaa actttaaaaa gtagtagcat gtatcactga ctaaaggaaa
59101 tggagaagaa gagagaaaaa gagcactgac catggagttg taaggaaWtc aaaaacaaat
59161 tcatccattt cacaattcac aattctgtct aggctgaccc tgaccttcat actcaaagca
59221 acacatgcac cattagcgtt tctcattttt taaatatatc caagcaaata atgcccatgg
59281 agatgScacc aacctgatct gtttgtggct tcacttggtt ctacaacaaa ctgtagagag
59341 gcttctttct tctttgaaga aactggaact gactttggat gtgatttctg gcactgagca
59401 cagaagaaat acaacattag aactatttta ttaacttacc aattcaatga aatatatttg
59461 ttttttaaat ggggggaaaa acaagtattc taaaacaaca aaccagctgt taagagtttt
59521 atactacttt ttattttatt tctaaagtaa tagctacttt ggaaaacaat ctggcattat
59581 tcagtaaagc tgaaatttca tacaacctat cacagaaaca atattactct aaacagcaac
59641 tatgtggaca gaagactgaa taaacttatt aaataaataa attatagtat agatatacca
59701 tattacatac gtaacaatgg aacaaataaa ctacaataat ataaattatt attttaacca
59761 gaaaaaaaag caacctgcag aagtacataa caatcttatc tatgtaaagt tcaaaataga
59821 cacgtaaaac actgtagtat tgagattcac tctgcaggaa agagttaaca taggaggcct
59881 gatatattta aaggaccagc ttacagggct acccccttggt tggcatctgg agacttaact
59941 tttagaatgt tccctccatt accaacagag aaggttgcac ctgctgcgcc agcctgccta
60001 gactgtttgt ataaataaac actgtgatct gtggtgaaca tctgctttct ttccttctgg
60061 gagtgcgtaa aattttggta ggcaggggat gcctatggga ccaaccccca gtgaaaacag
60121 gctgagtttc aaaaggcttc tctgaggaga aaaaatgtac acacgttact gcatttcact
60181 gctgaaggga atgagtacat tctatatggc cccttgRgga acacaggaag cctatatatg
60241 gattcctcca gactctgcta atgcgcattt ttcccttact gacccagctg tgtatcctgg
60301 ctgaatgact ataataagta tcagtcatga aaaacaccaa aagcaatggc aacaaaagtg
60361 aaaattgaca aatgggatct aattaaacta aagagctttt gcacagcaaa aKaaactacc
60421 atcagagtga acaggcaacc tacagaatgg aagaaaattt ttacaatcta cccatctgac
60481 aaagggctaa tatccagaat ctacaaagaa cttaaataaa tttacaagaa aaaaatcaaa
60541 cagccccatc aaaaagtggg caaaggatat gaacagacac ttctcaaaag aagacatcta
60601 tgcagccaac ggacacatga aaaaatgctc atcagcactg gccatcagag aaatgcaaat
60661 caaaaccata atgagacact atttcacaat agttagaatg gcaatcatta aaaagtcagg
60721 aaacaacagg tgctggagag gatgtggaga aaYaggaaca cttttacact gttggtggga
60781 ctgtaaacta Kttcaaccat tgtggaagac agtgtggcaa ttgctcaggg atctagaact
60841 agaaatacca tttgacccag ccatcccatt actgggcata catccaaggg attacaaatc
60901 atgctgctat aaagacacat gcacacgtat gtttattgca gcactattca caatagcaaa
60961 gacttcgaac caacccaaat gtccatcaat gatagagtgg attaagaaaa tgtggcacat
61021 atacaccatg gaatactatg cagccataaa aaaggatgag ttcatgtcct ttatagggac
61081 atggatgaag ctggaaacca tcattctgag caaactattg cgaggacaga aaaccaaaca
61141 tcgcatgttc tcactcatag atgggaattg aacaatgaga acacatggtc acagggtggg
61201 gaacatcaca caccagggcc tgctgggggg taggggggagg ggggaaggat agcattagga
61261 gatacaccta atgtaaatga cgagttaacg ggtgcagcac accaacatgg cacatgtata

FIGURE 4-Q

```
61321  catacgtaac aaacctgcac gttgtgcaca tgtaccctaa aatttaaagt aaaataataa
61381  tttttaaaaa agcaaataaa aataagtatc agtcatgagt tcaactatat gtcaaatctc
61441  ccgagtcctt ttagtgaagc accaaatgtg gatggtagtg ggacctctgt cacaagtggt
61501  agcagaagtg tggacacagt atcRgcacag tagagaaata aataaatgac agacttacat
61561  ggtttgcctg tgtccccaac caaatctcaa tttgaattgt atctcccaga atttccctgt
61621  gttgtggaag gcgcccaggg ggagacaatt gaatcatcag ggtcagtctt tcccatgcta
61681  ttctcgtgat agtgaataag tctcacaaga tctcatgggt ttatcagggg ttaccacttt
61741  tgcttcttcc tcatttttct cttgccacca ccatggaaga agtgcctttc gccctatgcc
61801  atgattatga ggcctccaag ccatgtggaa ctgtaagtca aattaaacct ccttttcttc
61861  ccagtcttag gaatgttttt atcagcagtg tgaaaatgga ctaatacact acactgctgt
61921  tgagtgacca cagtatggaa ttcaatgcca taaaaagaca gaggcaggct ttcattttat
61981  ctctgctctc tcttaacatc tcattaaaga tttttaaaaa caaactcaga aggatgaaga
62041  gagagacaaa ggagtagatg agacatgtca gcaaacattt ttaggttgaa aagcaaacta
62101  ctggtaattc agcaacccag agagaatgga aacatgctag caatggagaa aacccagaag
62161  cagacagatt actgtttatg gagaaaccta taaagattag tagaattgca agacaccact
62221  aaaaaaaagg ggcaggccag atgcagtggc tcacacctgt aatcccagca ctttgggagg
62281  ccaaggcagg agtatcactt gaggtcaaga attcaagatc agcctgggaa acttagcaag
62341  accacatctc tgtgtaaaaa taaaaaatta gccaagagtg gtggtgcatg cttgtgttcc
62401  tacctactca ggaggctgag gcaggaggat cacttgagcc cagaggtttg aggttgcagt
62461  gggctatgag catgccactg cactccagca acagagcaag accctgtctc aaaaaaaaga
62521  agaggcatgc aaacagaggg tgYtaaaagc gtgtatagtt gaatgtgtct tttgtcgcag
62581  aaaaaaaatt gtgtatagca agtagttaaa ctggctgggc gcagtggctc acacctgtaa
62641  tccccatact ttgggaggct gaggcgggtg gatcacctga ggtcaggagt tcaagaccag
62701  cctgatcaat atggtgaaac cttgtctcta ctaaaaatac aaaaattaac caggaatggt
62761  ggcaggcacc tgtagtccca gctgctcgag aggctgagac aggagaattg cttgaacccg
62821  ggagRMagag attgcagtga gccgaaatcg tgccacYgca ctccagccca ggagtgcttt
62881  ttctcaaaaa aaagaaggta gtaaaaccga agagcccctc ccaaacctat cagaagatta
62941  tttactctct tgagtttctt aatcagagag actgtgatat tagtggcata agacacaggt
63001  gaagttaagg gcaatacatg aaaaccagag atattaagta ggagcattta aactgcactt
63061  tccccttttgc cacatctgac tgaaattgaa aacaatcgca gttagacttg aaccgccaag
63121  cgggaaagat ttctttgtca gaaatactta ctaggccaag agaaaagaaa tctacagata
63181  ttaacaccag gataccaatg atgaaatgga ccaaactaca ccacagtgac atccaccagt
63241  ctacaaattt tactaggcat atagagcttc caatgtgttt tagtttccca ctttcaaaca
63301  agaggagaca gctaaggatc actggacata tgaaaaaggc gaggtaatga gcaaagtaag
63361  taaaaaaggt acttgcagaa aacagaaact gaagaaataY ttatgaaaag ggttagatga
63421  acctaatgtc aaggaatata ataaaaattt agatcactaa agacagatga catgatccta
63481  aaagcctgca aagaggaaaa aaatagtcac aaaggatcat ggaatcacaa cagtattctc
63541  aacaggagaa gctaaaataa aagaaagcaa catttgtaag tctaagggaa aatacacctg
63601  actaaaaggg aggcttgata tgaaaaaata tatattattt tttaattaaa ggaaaataat
63661  ttctaatcta gaattctgta ctcacccaaa ttacccatca tgaaggtaaa accatggtat
63721  ttttaaacaa ttaaggtctc aaattctcag aaagctacta aaggatgagt cccaccaaaa
63781  aaggagaggc caagaaaaaa ggaagatttc agatccaaga aataagagtt gcaattcagg
63841  agaaaagcaa agggactctc tgtgatacag agagagtaca tgacaaaacc tatgtgatag
63901  gtctaagggt agattgaaga aaaaaaaaag agcagggaca gcaaagctga aaatgataaW
63961  ttaatgtgtc tatgcatact aagacattta tacttagacg tttgtcatag aaaactggtg
64021  atgaattagt cacaggtgca cagaaagtca aaaatggaaa aaccaaaatt aaatcaagag
64081  aaaattaaga actgtacata aagaaaattt aaccattagt actctacata gctgtgacta
64141  ctgaaacagt cataataata ccataaacac taactactaa tcctactaaa aattatgatg
64201  tactaaagaa tggaagttta tgtgttgtgg tcggaagtag aaactatatg acaaataaag
64261  tctaaacttc catagtagaa agccaataaa taatagcaaa gactggRaaa aaaaatcaag
64321  aaacagcaat ataaacatta tttagaaata tggagtcaaa atacctagag taatagctaa
64381  gactaaaaag ttggacttgg aggaggaaaa gtggggtcta ctatttttt gtggaattta
64441  tggaactact ttaacttcca tgtacataat gcctttgagc tttaaaaatc aatcttccaa
64501  caaattaaga agaaaagtat gtaaaaagat agggaagaRa ataaaatagt gaaccagaat
64561  agactgctga ttctttgaaa atgttaatga cagaaaaata cctgtaacaa cacagataat
64621  ttaaaaaggt gaaaaccagc aataaaaaga gagtaataac tatgaggaag atcatgaagt
64681  tgactataca caaattgaga acagaggaaa tgtttcaaga aaattataaa atgtcaaaac
64741  tggcaaaaga aatataaaaa atgcaataga tcattgaata aagtggtaat cacacacatc
64801  cccaccaaaa cagctccaac catatgtttt acaaatgatc caaatttgaa gaaacagatc
64861  atttttgctt aatcaagtta tcccagaaaa cagaagagaa aaactaattt atttcatgaa
64921  gctcggttgt gtttcagtta aaacctacgg agtcatatct tattttactt aaagacaaac
64981  aaatatgaag taataactag acctaggaat ctcaagttcg caaacagaaa agcctataag
65041  atagataagt cttaagtctg attttatttt caaatcaagt aaacaaattt cattccttat
65101  gtggaaggcc taggccaagc ttgctgcatg gaaaggccct acctgaaaaa aaaaccctaa
```

FIGURE 4-R

65161 cacctacaaa gtctctgttg ttttgaagca tccatctcaa ggattagaaa aaattaataa
65221 tataaaaata aaaataaaga atatatgaaa aataaaatta agaaaaaaat aatgaaacta
65281 aaaacaagca aatgagagga ccgacaaagc caaacttggt tttctgaaag accaaaattg
65341 ataaaccctg acaaaattaa tcgatgaaaa agacaaggca tagcaggcac agtggcatgt
65401 gacttacgag cagtaatccc agctattcct tagcttaggg aggatcactt aagcccagga
65461 gttctaatcc agcttgggca acataccaag atccctacct ctaattaaaa aaaaaaaaaa
65521 aaaagataag gcataaataa ccaatgtcag gaataaaaaa ggggatataa ctgtagatcc
65581 tacagacatt aacaagactg gaaaaagaca tgacaaaaca acattatcca aataaactgg
65641 aaaatttaaa taaaatgaac aaatccctaa aaaaagaaaa gtagaacatt taaaaaaaat
65701 agaaaatcta aattgtctac tgaaactRtt aagtaaaatc taaatcttta atgttctata
65761 tcatcatcta ggtagtagtt gtaaaaattc aatgaaatat acacgcaatt tgtgtatttt
65821 aatgtattat gtttcataag aaaacaaaag tgaggccagg cacagtgctc atgcctgtaa
65881 tcccaatgct ttgggaagct gaggcaggag gattgcttga gcccagaagt ttgagacaag
65941 cctaggcaac acagtgagac cctgcctctg taaagaaaaa ttttaaaatg gccgggtgcR
66001 gtggctcaca cctgtaatcc cagcactgtg ggaggccaat gcaggagtat cacttgaggt
66061 caggagttta agaccagcct ggccaacatg gttaaacccc atctctacta aaaatacaaa
66121 tattagccag gtgtggagat gggcgcctgt aaccccagct acttggggga ctgaggcagg
66181 agaatcattt gtaccccgga ggcagaggtt gcagtgagcc aagatagcac cactacactc
66241 cagcatgggt gacagagcaa gactttgtct caagacagat agatagatag agaaataata
66301 aataaataaa taagcaggcc agcatggtga tatgcaccta tagtcctagg tagtcagcag
66361 gctgagacgg aagggtttga gcccaggaaa agttcaaggc ttcagagagc tgtgatcatg
66421 tcatggcact ccatcctggg ctaaagagtg aaaccctgtc tcagaaaata aataaataaa
66481 taaataaatt agtattttaa aattctcgca aagagaaaac Kcctatgaat tccatcaaat
66541 atttgagaaa gaaacaatag tggttcataa aaacacttca ggagaaaaga gacagagaaa
66601 atactgaaca actcatttta caagaccagt acaaacacag agtagaacct ttattttgaa
66661 catttcttat ttattataca tttaaagtta tgaattccct ctaacactat tttagctaaa
66721 tctcacaagt tctgataagt catattatta tcttcattca gaatatttta tattcccaac
66781 atgctttctt ctttggccta aagtcattta gaagtcattg cagtggacct tggctttaga
66841 gcacaatgtc ttaaaaggga ccagatttac cctcctgcct gaatcaattt taaaaaaaga
66901 aaaaacatac aagattctca agaaatacga cattgaaaaa ccaaagagaa tgattcttga
66961 gatgggaaac aaatgaggtg aaccctctga tcacccattt ccagatgtga cacacagggg
67021 aaatcctcag ttgaggagcc aaagctgagg attcaggaag accaaagtgg ttagaatttg
67081 caggataaag tactggacag gagacggtta catagagaaa actgcagaat gtgcagatct
67141 ctctcaagca ttcaatcaaa tattgatcag agaatgcgtt gtaagaaaac cagagaaaga
67201 accacccgaa gccttgaggg ggaatatcct cccagttaca aaggacctag aataatctac
67261 cagccagagt ggaaaacctc cataagtaac gggacacagt agagaagaat tttgcctcag
67321 aagggaagaa aattagccac agactatatt ttgctagtct ttattgtttc cacagcttct
67381 catatatcta aaatttgttt ccatttttta aaatagctgc tttagcagga gtgttagcct
67441 tctgtgactt actgcattct acctagaaga ggaacctcta gcataacctt gataccaaag
67501 ccctgtaaga acatttttca aaaaaagaca atattattca taaacataga tgcaaacatc
67561 ttcaatatta ggcaaaccta atcaaaaaaa gttaatactt tatgaccaag tcaagtttat
67621 cccagaaaac ctaaagttgc tttaactttc aaaaaccatt gtaatttatc aagtttcaga
67681 aggcctctca attctgaaag aaatggtttt gtctgcattg cttggaacca aaaggctaca
67741 tatgtttaat atgttgattg ccttttacag ttttctagta agtgttactt ctagcactat
67801 atgatttttc taattagtat tgtgcttttg aataatttgc cttagattag cttcaaatac
67861 tttatttgaa agtctgctac ttagtatagt taagtcaacc aatactttat ttctgatcaa
67921 actaacagca attaaaatat ataatcaaca gaagattgat ttccccgtat aacaatgaag
67981 tgccttcatt tattagaata taaccaaccc tgccaacctg attagcaact ttgatcactg
68041 cccatttttcc taaaaaccta atttgagtat tattttagga gaagtgaaaa atgtttatta
68101 taatttgata attctaattc ccataatccc aattccaaca gaaaatatca aataaaactt
68161 ttcacaaaat acctccaaat tttaYagaaa gtcaatggga atatagaatt ccctttcaag
68221 aataagactt acaaacaact aaaatataca tgagtatatt gtgtatacct tcattcatta
68281 aaatattaaa cctataccag ctcttcttta tcacttacct ctttgcttgg tgactgaata
68341 caagtgtctt ttattttgcg tgttgaatta ggcactgatt ttgtagaatt tgtactaaaa
68401 tcattggcaa gacctaaaac aaaagtatga aattttcaaa aacaagatat agtgtaatat
68461 ctattagtca caattaaaag tagctgtata tgacatgtga agaaccatct gttgactcac
68521 tgaccacaat tagcttaaat caagtagaga cctcaattta atattcaaat gttcctgaga
68581 agcatcagga aatgagaaaa gcaaacaaaa cacctgaagt atcaatatat gccagctctt
68641 ctttcagagc ttgtatagaa aacattcttt tctatcatct gcccctattt ctttggctgg
68701 cttattctct tattccaata cctgctgatg ctaatcagca cttggaaaag tttttgctgt
68761 aatttatgac ttccttaaac agatatgtaa atcaccaaag attatctcaa attctccagc
68821 gattccaaga attctacaca tctattttca ttatatattt tttctggcta gatgcagtgg
68881 cttatgcctg taatcccaac actttgggag gctgaagagg gcagatcact ggagggcagg
68941 agttcgagac cagactagcc aacatggtga aacctggtct ctactaaaaa atacaaaaat

FIGURE 4-S

```
69001  tagccaggcg tggtggcaca cacctgtaat cccagctact tggaaggctg aggcacaaga
69061  atcacttgaa cctgggaggc aaaggttgca gtgagttgag atagcaccac tgtactacag
69121  cttgggtgac accgcaggac tatgtctcaa taaataaata aataataaaa tacttttttct
69181  tgtgtatgat aattatgtat ggacttccag attacagtaa tttaatgtga actaaacggg
69241  atgacaaatt attgtcaaat tattattttt tgtcagtatt atcctaagta taccagaagg
69301  tacacaaagg attgcatatg aactcaaaat accagggaaa atattttcca tatatttaaa
69361  ccctagaaca tatttgctgc atttgatatt aattaccatg acaaattcat attttcatcc
69421  atctaaccct taaattccat cttgaaatga agccttactt ttttcttcaa aacagtcttg
69481  taagatttcc agaacattct ggccttgctc tgtgttaatg tcacgtgccc taataaagga
69541  aaaatacagc ctgtcatttt cagaagatat ttctttatat taccatctct ttgaaaattc
69601  cacatctcat tataaaatat aaatttactt aatatgtaaa atRtctttac tatttgaaga
69661  caaaacaagt acttgagggt ctatacattc cctcaagtga actctttcct tcctaataag
69721  aattagtata tcataatcta ccttaacaat taatatttct tagttgaaca gccatgtcta
69781  aagctgtttt agattactgc tgtatctcaa aaacacaaat ttttttgaat gaatccaatc
69841  agtcctgaaa tgcttcttta acagacttta cagtcacatc tccaagtaac atattcctcc
69901  ctataaattc tcaaatgcat ttctaattca gtatcattga ttaatcttaa cagtttatga
69961  atcagtataa attttaatta aaaattatat tgagcttgta gattcacagc cagctataca
70021  acataataca gacccccttgt tcactttacc tagttctcct cagtgctaat attttgcaaa
70081  actatagcat attatcacaa tcaggatact gatattgata caatctacca atgctattgt
70141  gatttctcca gttctacatg tactcatttc tgaacatact catgtttgtg tgtatatata
70201  caaaatgtat cacatctgta ggtgcagata aaacaaccag agtcaagata ctgaatagtt
70261  ccatcatcac aaagatcctc tattgtcctt ttgtaatcat aaccaccttt ctctcatctc
70321  cttcctgcct cctctcattc ctaacctctg gcaaccacta atctgtcctc catttctaaa
70381  actctgtcat gtggaaacaa ttatatgtaa atagaatcat acaacatgta tcctcttgcc
70441  ttggcttttt tcactcaaca caattctatg gaaataaatc caagttgaat gtaccaatat
70501  ttcattgctt tttagagtca agtaatgttc catgctatgt gtgtaccaca gtgtttaacc
70561  atttacctgt tgaagaacat ctggatttca atcagtataa aattgacatg catctctaac
70621  agagtggaca aacttttgct taaagggcca gatggtaaat attttaggct tgtaaaccat
70681  agtctctgtc acaaccagtc gactctccca ttctagccca aaagcaacca cagacagtat
70741  gtaaacaaag aggcatggct gtgttccaat aaaattgaaa taaaacctgt gacaaaacat
70801  ttggctggcc tatcagccag agtttactga tccctgatct acaacattcc tattattgag
70861  agctggaaaa gtactatata tgcagtaatt aaaaactgtg agccacagtg cattatcagg
70921  tcatccatca taaagtttta taactgaaag atctcagaaa tcatccagtc caataatcat
70981  catccaaata gatgtgtaaa gttacttttg gagctttgca gaaatacact taattggtca
71041  cttcccacaa accaatttaa ttagatttcc tagggtgatg gaggtaattc tgatgtgcaa
71101  atctagttaa aaacccctag ctgaatccta tgccctcatt ttgcagatag gcaaactgaa
71161  acccaaaatg gtcaaatggc ctacctcagt ttgtaaacct aagtagtatt agtggccggt
71221  tttctaccag aatatatata ggtgactatt ataaattctg gccttgacat caccccaaaa
71281  taccacctac aggtgcacat attgcacagc tataaataca ctcctgctat tacctgtttc
71341  aaggatgaga tccaaaaaga tcagatttgc caaggtgata atgcagagtc actattagaa
71401  cgcaggtcac attaggcaac agcttttca aaatttatca aattcgtctt ttcttcccct
71461  ttgtcattat taccttagca taaaagcacc tccaccatgt gtagattact acaacaccct
71521  ccttaagcta ctttgtaaac atcaaacatc tttgcttttt taaatgtgtg tgtgtgttta
71581  tatcagatag cccttattcc acagaagatt taggagagct tattcaagcc tataatagat
71641  gcctaaagcc taacatatcc attttatggc tcaatccaac tcatatggtc ttgttttcca
71701  cttctaccca acatattcaa ccttcttaaa tcaggtctct ctgctactga caaatatatg
71761  tattcctact ttggttcctt cgttaagtta ctttctatgt ctggaatata ctttatctta
71821  gtctttaaat acctgttcaa ggtccaccaa agttacacta caactttaaa gcttttctta
71881  atatcaacaa cacacaaatc cctaatttct cacatggctg tctatactgc aagttaatgc
71941  ctaattatat ataggcctac tctctatgat acacaagtca ttacttcagt attgcaagta
72001  ataatttttg tccattaaat tgtttcacaa atgtacatat acattttccc aacagaaatc
72061  taagcttctt gtaaacaaaa gaacatactt ctgtgcccca ttggacctac taaaatcaca
72121  gctcaataaa catgacgtta ttcataataa catcttcact tttaaaaaca tgcttttcat
72181  cctattacta gtaaatattt aactagaaaa tttacttttt acaggtttta caactaaaag
72241  aagaaccaaa acaacccaaa caagcactca acaaatgcac gacagcatgt actatttttc
72301  tttttgcccc actgaataag ctacatcttt ctaattttac catttgattc attctacaaa
72361  atgcttttcc catggcagtc acttagtcag aactgagact ccctcttcct ggttccacta
72421  ttttatcagg tgaatttctt ttttcacagc cacaaagtca ttgcactgct tgcattaaca
72481  ttcccgttag tataagatga ctacccttaa tcagtgggca cagaaatcat tatattagtc
72541  agaggtgata gtttttcctt tcttcattta aaaatctatt aacgatgaga ataaaaataa
72601  tcttaaaagc ttaagtacgt aaatcacctg gaaggtcgac aaaatcttct tctgtagcca
72661  tttttgagat gatcctgaaa gaaaagtaaa tcatcaattt ggtttcttta agatagagtc
72721  gagtaaaatt tattttcaaa gtaacttcat ccagtatggg catgctatac accaaacatc
72781  tcctaagggc tcaaaacaat tttgaggtta tgttatgaca taacataggt tatgacaggc
```

FIGURE 4-T

```
72841  atttaaaaac tgaactgcag gcatctggta agctttattc cgcttctctc cccacagtga
72901  gaatggcttt attgctttct tctgattata tataatgtaa acaatatatg tgctcatagt
72961  aaaaagtata gattatatta ttctttcata caactcttta tgtttatatg tataacgtca
73021  cataatggga gatcatcgaa gctgaagatt ttcttagaag gggagataaa ttagaaaaga
73081  aatgatacgg atttttatat agtaaccata accaaagtat gaaatataga atagtgaaca
73141  ccaaagaagt ccagaaatat gtaaacaaaa tgagttgtcg cccccattag tcattcacag
73201  cttgtacctt ttaactgata gctgtctttc agaagagttt gtcaataaca tatgctccat
73261  aaatcagcct caagaactgc ataaaaggcc accattttga cagaagaact gaaaacgtca
73321  aattatttcg agacgtactt gttacataat ctgtttctcc caactaccaa gtgtgacgca
73381  aagacacact tttaaggcgt ttcaatactt ctaccatggg gagaaaagtt atttggaggc
73441  ctaggaactg gagctagact aaaaagcaga agccaagatc gcgccactgc actccagcct
73501  gggggacaga gcaagacccc gtctcaaaaa caaaaacaaa aaaaaaccca gcagaaatca
73561  caagcacctc tcaagagtaa gacgtaggca cagacgaaat gacgacatga aaacttaaga
73621  aataataaaa gtacagtaaa ttaactctgg ttatcgacac acgatcacac ttgatttcaa
73681  aaattcctgc ggctttccca ctgaaatccc tcagagatca cagcctcagc ctagcatttc
73741  cttctcccca gcctcgggcg cccgtgcccc agccagccgc tcaaccactc gcctggagcg
73801  gggggcctgc acttaccaga ccggacgcag ccatgttccg gccccgctga gccagcgcaa
73861  ctgtctgagg tggaagccca cacggaccac agctccagga agccgagcaa gaaacgaatc
73921  gccggaatac caggccgcgg ccaagcaata accttaagtc tcaggcgact gccgcgagag
73981  ctgcgatccg gaaggcgcgc gctaacaaca ctcccccgcc attggctccg ccctcaacaa
74041  acaagaacgc ttggcctgtg cctgggtttg aacccggaga cacatgcgca aaacgggccg
74101  cgccgtccaa taggaagcaa gccgttagtg ggctcgcccc ttcatgctcc aggccaacgg
74161  ccctccagcc tgcagaccca cccgcccgcc ctacctctag ccgctctgct cttcccacct
74221  cccttcctca gcgccgcgcc cgggacaggc gcgagactgg gcttttatta ggaaggcacg
74281  atcctgggaa ccgctccttg tgcccaatgt aaaacttctc aacagttcta ttaaggaata
74341  ataagtaata cttggaagga tgagaggtag aaattgcaga cgacttttt tttaatagaa
74401  tctttttttg aagagatttt tctaatgcag attcgtcagc cacgcccagc ccatctaacc
74461  atcatctctg aaaataattg aataattatg tataataatg gtgatgatgg cttgtattct
74521  ttttcaagta tgtgggcgat atttttatga tgcattgcca acaaatgacc cacgtaaata
74581  tctatcattt aagtatgatg cactgaaaag aattctatac agaatcctga aacctgccag
74641  ccattgacta gctatgttat ggagggcaaa attgcggcac cattctgggt gtttccatag
74701  tgatagaatg gaagtctaaa ttagacgctc tgaagacttc tcacctgacc agtgtatgat
74761  tctatgaagc tgtatcttat tttattattt aatttaaaat tacaagaaaa ttctaatcag
74821  aaaatagttt actaattgaa taaatacatt aatcaggagt tgagctaata gatttttaaa
74881  acatgtagtg ctttaacaaa tgctgtcaMc tatgaaggcc aatttttta ttttatttat
74941  ttttttgata gaaaaaaagg aactatgatc tctgactgaa attaggcctg gattgcactc
75001  cattccatta attctgccat tatgggcaac taagttatct gttccttaat tctcttgtct
75061  gcaaaatgag aatactatta aagtggttga gatgttgtaa ttgaataatt cctgttaaga
75121  tctaagcaca atgtttggaa catagtaaga attcaataat aaagctattg gaatcattat
75181  ggagagtttt ttactatgtt tggcagtttt gccaatcagt ccaacttata atttggtgct
75241  gtcacatttg gaatagataa tgataaacac taacatttat tgactgcata ctctatgccc
75301  agttctgaaa tgctttgcat gcattgaaga gaacaattaa tattcctgtt ttataaacaa
75361  gggcactgaa acacagagag ctaagtgtct tacccaagat cacatcacca ctaaatggta
75421  ggatgggtta ttatcccaag aattctcact ccacaactga gctcttaatc actacactct
75481  ttgcctctgg gtaagaaatt ttagcaagca cgattataga tgtaaagtcc cacatgtaaa
75541  attatatcct cctgcttcca ctggaaacta agtcccacac ttattttgag aaatttggaa
75601  tgcaatggct tgtgctgctg cagaaagcag gccacagagg gcgttgagga ttaactgttc
75661  taaagaaaat agagcagtac tcttataaaa acagttgtgg agactccaca gtttccagag
75721  aaacgtttta cttccatatg tacttgttgt agtacttaat agcccaccat ggttttagta
75781  tcctaaaata ggattgctaa ttaatggggg gagtaatatt tatttaaaaa cccaagtttg
75841  tcgttataaa tctccccccа aaatacatca aagattactt cttatgaact gctaggacaa
75901  tcttgaagtt ttctttcaca actacttacc tagttaaggg gtgtgtgtgc atgtgtgtat
75961  ttacatatca gtatatggaa gtattaacat atacataatt agatgtgtat atgggaacat
76021  aaagtatcca gaaatttggc ttgaaagaga agtaaaaaaa acttttcgtt acttatggtt
76081  ttaggtacat aatttaaaat ttgtaaatct tgaaactatt tcaaagaata ttttgtttta
76141  gatctttgat ttccatcaaa taactatggt atgggtattt actaccaaag tgtcttccat
76201  taagtgttag cacaaaaata ttcacaactg aactaacact taaaaattag tctgtattca
76261  ttcccacccc acagcttaac atgtgcatct gtcctggtga gtggataaac aaatgtacct
76321  caatgcagtg caatactctt cagcaataaa agcaaactac tgatatacaa tatggatgaa
76381  tcttgaaagc actgtactaa gtgaaagaag ctggacatca atgactgtat gaaaagaaat
76441  caaatcgctg gttaccaggg gttgaggagg gtgggggaaa aaagactgta aaggaataag
76501  agggaacttt ttgcggtgct gaaaatattc tatgtcttga ttacgatggt agttacatgg
76561  ctatatctat ttgtcaaaat tcctagttag aactgtgccc ctgaaaaagg cgacctttt
76621  ttttttgaga aaggtcttgc ttggttgccc aggctggagt gcagtggcac aatcaaggcc
```

FIGURE 4-U

```
76681  cactgcagcc tctacctgct gggctaaaat gattctcctg cctcagcctc gtgaatatct
76741  gggaccacag gcatgcacca ccacaccctg ctaatttttt ttttttaat ttttgtaga
76801  gatgggatct tcctatgttg cccaagctgg tcctgaactc ttgggctcaa gtgatcctct
76861  ggcttctgcc tcccgaattg ctgggttaca ggcatgagcc accacacctg gccaaaatag
76921  tgacttttac agtaaattat acctcagtaa atctgtgaga gcgatgataa aagagagaaa
76981  gagaaaatga gagagactca aaggcactgt atgaaaacca aatgcaatgt gtaaactttg
77041  attagagtct agttcaaaac aaagtaactg taaagagcat taagcaataa gtgaaaatga
77101  aaaaattaat tttcagatga tattagagaa ttgtaaagtt ttcttaggta ttataatgtg
77161  attatgtcta tgtaatttgt ccttactctt aggatatgca tatcaaagta tttagaggta
77221  aaatgtcttg tgatatttgt aagtactttg aaattgttaa tcataaagag attgagatag
77281  ttaaatacat acggcaaaat attaacaatt attgaatcta gttcattggt gtgtttatta
77341  acaatactct ttcagtttct ctgtatgttt gaaatatttc atagaaaaat gatggggaaa
77401  cttgaactta ataaaattgc ttaagttcta aaaacttttg caaaaacatt tattatttat
77461  tctgcaaaag tcaaaagtga aaatattagc aaatgactat catataattt tgtcagaaaa
77521  attacacatg aaaaatgaat ttcttctaaa aagtatagca tagtcacaat tataacaata
77581  tttgctaaga taaaatagac ttagttttat gattttggtt ttttttata atgtaatttt
77641  ttttctttta ctccttactg ccaaatgttg gtccctctgt taaggagatt catttaagta
77701  tccttttctg ataaaaaatt atcctccaat agaaatgcta tctatgctat gtagaccata
77761  atggaaaagt taataaaatg tgcttgcaat gtacaggtag gtggcagcaa tgtttacaaa
77821  atagttttac cacgctatga gttgttaagg tctgcaatgt gtcattaatt gtaaagtgct
77881  atataaatta actgatcttt acatttccta aattaccttt gtgtcaacca ttgcttatta
77941  gttatagtta tattctatta tgtacacatt gtttctctta agatcacaac aaataacact
78001  attctgtaga tctccattct gcatcctata agataggata gtttacttct agtcaactta
78061  ttttctactc acacaaagga aattttgtca ttgaaaacac attaagattt tcaaaagctt
78121  aataaaatgg ttctgcattg taattttact cttgagagga cttaaattac cctgagtgaa
78181  aattggcaag taatgtagga tactgttttt aaatttcagg cagtcattgc tttgcaaagt
78241  agtacaggac ccttaaaaag accctgtaag ctgaaaccat gcaaaccgat ctttgcttgg
78301  gaaaaattct gattgtttca tgacctttaa aaaaaaatta tcagctggag gcggtggcta
78361  acgcctgtaa tcctagcact ttgggagact gaggctggcg gatcacctga ggtcaggagt
78421  tcaagatcag cctggccaac gtggcaaaac cccgtctcta ctaaaaatac aaaaattagc
78481  tgggcgcagt ggctcacgcc tgtaatccca gcactttggg aggccgaggc aggcggatca
78541  cgaggtcagg agatcgagac cgtcctagct aacatggtga aaccctgtct ctactaaaaa
78601  atacaaaaca attagccgag catggtggcg ggtgcctgta gtcccagcta ctcggggggc
78661  tgaggcagga gaatggcatg aacccgggag gcggagcttg cagtgagccg agatcgtgca
78721  actgcactcc agcctgggca acagagcgag actccgtctc aaaaacaata aaaataaaga
78781  tacaaaaatt agccaggagt ggtggtgtac gcctgtaatc ccagctactc gggaggctga
78841  ggcaggagaa ttgcttgaac ctgggagaga ggctgcagtg agccgagatc acgccactgt
78901  actgcagcct gggagacaga gcgagactca gcctcaaaaa aaaaaaaaaa aaaaaattat
78961  caaaatatta acatgtctct tgctgtcagt tataaattta tggggaaatg aaaaaatagt
79021  aaaactaata cttatatagc acattgtaac tcaaagcatt agaaacattg agcattaaag
79081  tgctttagtt ctcagttttg ttttctattt ttttatttt ggctcctgac aggaaacatg
79141  ctttagttct ttgtaaaaaa aaaaaaaaag aaagctatca agaatagttt gaacagtgcc
79201  tgcctcctca gctgtaagtt atgatacagt gtgagcatct tttggtgaat tgttatacta
79261  cttgctaagc ctgggtcatt ttaaattaag tatatgtggc ataggatata taacatattt
79321  gtaaattcaa gtgtctatgg aaactcataa acaataaaat accaaaatgg taacttgcag
79381  acagtaccta aaagtaagag ttgttttgtt ttgttttgtt ttgttttgtt ttgttttgtt
79441  tttttccgga ggctccatca ttcacctaaa agtaagagac aacaggaaga taaactagaa
79501  gcaaagacac agggcagaag tgatgtgctg ttgacctaga gaaaatctga ctaatattga
79561  aaagtgcggc caggcacggt ggctcatgtc tataatccca gcattttgag aggccaaggc
79621  agcaggatca cttgaggcca ggagttcaca accagcctgg gcaacatagt gagacccagt
79681  ctcaacaaaa aatcaaaaat ttaaaaattt agccaggtgt gattgcatgt gactagtttc
79741  agcttctcag gaggctgagg tgggaggatt gcttgcccca tggccctctg cctaggcaat
79801  agagcaagac tccttctctt aaactaaaaa taaaaatttt aaaagtgttc atgttgacag
79861  aatggaacta gattatgggg caacttcaat gtcaggataa cttttcactt atcctatgga
79921  caatggggaa ctgtggtata atcagggttc aattacataa aagaaaattt gtagccaggc
79981  gcggtggctc acgcctgtaa tcccagcact ttgggaggct gaggcgggcg gatcacctga
80041  ggtccggagt ttgagacccg cctgaccaac gtggcgaaac gccatctcta ctaaaaatac
80101  aaaaacaatt agccgggtgt ggtggcacat gcctgtaatc ccagctactc gggaggctga
80161  ggcaggagaa ttgcttgaac ccaggagacg gaggctgcag taagccgaga tcacaccact
80221  gcactccaga cgaggcgaca gagcgagatt ccgtctcaaa aaaaaaaaaa agaacaattt
80281  gtgatggcta gttttggcaa aaagggatct gctacctgca tttaacagct cataaaattt
80341  ttgagaggat gaagaaacag tgtattttgg attttcagga ataattccca aatgcacatt
80401  gaagaactaa gccattaaga ggcagctgat gtttctacag ttataaaaca aatagaaaaa
80461  ttgggaagct gacattaatg cttatgtctc caagatcata ctgacatcat gcaatcagga
```

FIGURE 4-V

```
80521  agccaccaca attacaaagc tgccacgtta tggaaatcac tataagaatt aaagaaagag
80581  gaaagaaaca caaaaagtgg cttgacagcc aaagacaggt ttattttaga gaaaacccaa
80641  gggggcttct ggccaagtta ggtcagaggc acactctctt acagactaag agtttttaag
80701  gattcagggt gggagagttt atcagaggct tgtactgctt ctgtgtctct ttgctgtgct
80761  tatctgggag gaagacttgt gtgtctgttc ccatacatct ttctgcgctg caggcatact
80821  cccccgagtc tgcttttagc ttccctatct tagtgcacct gaagggaaag gaacgtgctt
80881  attaaggccc actgttttac tggggcccat tgtatgaggg tgaagtttgg cagttaccca
80941  agagactttc cccccacctc cctctgtgcc caagctgtct tatctgtttt actctctgct
81001  ctttctggct gcttgtagtt agaagagaag tgatttcctt gaaatgcatg aggctaaaaa
81061  gggagctgga acttaaagtg gcagtatttg agatgatggt gctcctgctc tgtcagtcac
81121  taccaataag aattctccaa tgagtctggg tatggtggct catgtttaat tccagcactg
81181  tgggaggcta ggataggcaa atttcttgag gccgggagtt ggagaccagc ctggccaata
81241  tggtgaaacc ccatctctac taaaaataca aaaattagca aggccatcat tgttacagaa
81301  gtgatattag tagctctgtc agaacagttt aatttgctaa cccttataac tgtcagtaca
81361  aagagttgca tctgggatga taagagcttt tcctattacg tattttcaaa ctatttttgt
81421  agtggaaaat gtcacttttg ttactgagga gtagtttttg gatgccaggt agatacggga
81481  aactggtttt catactaggt aactcacaga ataatggtag cttattctat acttatacag
81541  tagtggtggc acactactgt aatcccagct actcaggtgg ctgaggcatg agactcgctt
81601  gaactcggga ggcggaggtt acagtgagcc aagatggtgc cactgcactc cagcctggag
81661  gacacagcta gactctgtct caaaaaaaaa aaaaaaaaaa aaaaaaaaaa aagaatcctc
81721  tcagagcaga aagcctcctt actgcttccg gctccagagc cacaccttac ctaaaatgat
81781  atgatgatat gcacaagcaa actgatgtct gtgacttgac tatcaacacc catgaagcta
81841  gtgatcagac attgggactt ctgataactt tgccacagaa aaattagatg cctacaggcc
81901  atacatgcct ttttattggt tagaaaagcc aaaggttgcc tggtgcagtg gctcacatct
81961  gttaattcca gctactcaga aggctgaggc aggacgattg cttaagccca ggagtttgag
82021  accagcctgg gcaacatagc cagacctcct ctcaatgaag tttaaaaaac aaacaaacaa
82081  acaaacaaaa acaaaggcag cataagaatg aaccccgcct cccttctgcc ttgccaatca
82141  tgttgaatgc ttttcaatct tagctgcaaa gaagtctgaa aaatgtagat tttatctttt
82201  tagcctctgc aataaacaaa agcacactag aagggtttgg aatgacggta agcaccaatc
82261  caccttattc cctaaggaag gcatgaagga tgagaaataa aagagaaaac atagacttaa
82321  cctttgtata cattgagttt gaaatggtta aaactgaatt cattatccct ctaccaccca
82381  aactatctcc ccactctgaa tttctagttt ctattaaagc taccattatt ctgtgagtta
82441  cctagtgtga aaaccagtgt cctgtatcta cctggcatcc aaaaactatt tctcagtaac
82501  aaaagtgata ttttccactg taaaaatagt ttgaaaatac ataactagga aaagttttta
82561  tcatcccaga tgcaactctt tgtactgaaa attataaggg ttagcaaatt aaactgttct
82621  gacagagcta ctaatatcac ttctgtaaca atgattgccc acctttttt tctgtttcaa
82681  aagacttaag ggaaacattt tctgagtaaa aatggctcac tgtatcagta attctcaggc
82741  aaaggagtag gaagctccct ccatattgta tgtctgataa tctctggagt tgtggtatta
82801  caatgtgcat tttctccact gaaaatgctg catgtctggg tgtggtggct aatgcctgta
82861  aatcccagcg ctttgggagg ccaaggcgga cggatcattt aaggtaagga gttcaagacc
82921  agcctgacca acatggtgaa atctcatatc tactaaaaat acaaaaaatt agMcaggtgt
82981  ggtggtgagc acctgtagtc ccagctactc aggaggctga ggcgggagaa tggcatgaRc
83041  ctgggaggca gaggtggaaa tgagctgaga tcatgccact gcactccagt ctgggcgaca
83101  tagcaagact cagtctcaga aaaagaaaag aaaagaaaag aaaaagaaaa tgtggcttaa
83161  ctctatcaat agtaataaaa aatttactca taataaagat tgtattggat tcataccagc
83221  cattgataag ctataagtat ttcacacaat ggttaaaatt aaagttgtat tcttcaggta
83281  tgagttgcaa aaataaatgt attcaattaa aatatcttag aaatatacta ccgcaaatta
83341  taatgcatat ggcaataaat catcagtaag tattaagcaa aacaaaggtc agaaatgtgg
83401  aagtactgga gcagggttgg ttgatttgat tgattcagta ctaataaatg gatagaaaag
83461  tattttagag atcctcaatt attgtaaaaa atgtggattt ttactcagat gaccttataa
83521  catagaggag tcaatacagc actgagattt tggtttatat gagttaaatc ttatacttca
83581  cccaaagttg ttaagttata tttgccataa aacctgattt gtaacaatct tctgtgacac
83641  agatgctact gatggctgta atgttgggct ccaatagttt aatagtaaca tgattgttct
83701  acaaggcaac acaagaatca tatgtcaacc aatggtcaca tatatcagaa aaaaggaaac
83761  tcctgtcaca atggaagcta ataattctcc aaactagaat catatttttt agctagaagc
83821  tattcagtaa atatatgata ttgcaatcta cttataacca acccctactt caacacaatc
83881  tttcctttgg aaaaactgtt tcttgttgat actcaatgta tacaaataac tacaacagga
83941  caaatactat aagaaagaaa tcgtgcaaag tgctttggga aagcagaaat gagagtaata
84001  actttttttt ttgtgggggt gaagaaagct tcaggaaatt aggagatgtt cagaattaag
84061  gatagattca gcctgggtct tgaaggacga gtaggacatt accatataaa caaagagaag
84121  gaggggcatc taagcagagt acagtgactg aaaaggtatg aggagacagg gaagatcata
84181  gggtattcag gcaactgaga ctaaatgaaa ttaccagagc aaagggtatt ttgaggagtt
84241  attatttaaa aggggccctg ctataaagga ccaacgtgcc atgcttgagc ccattgttgg
84301  tattcaataa gtatttagtg ttcaataaaa aatatttaaa aaggaatctc aaaagcaatg
```

FIGURE 4-W

```
84361  agaatccttt aaagtctttt cagtagggga atatcttgat taggttgaat attaaatgaa
84421  tcaaagaaga aaaagagtgg aaactaggag actacttaaa agccttttgc agctcttcaa
84481  acaataagtc atgtagtcca tcgcctcctt gtataggtga ggaagctatg ataccagaga
84541  agtcaagtgg tttgttcaag gtctcaagag ctggttagta gcaaaattag tttatgccct
84601  atgtccaaac atattttcta ctatatttcc ttacctacaa atttattctt cactgcctaa
84661  gacttgaggt atctgtaaga ggatttccat tgtaatggtg cagtggtgca atgacactag
84721  tcattgcaga caggactaac ttgttaatat tgtttaagca agatttttgc attttgtagt
84781  ctttttcagg gagatataat ttgattcttg gtatattaag aatgaatttg acgaacaaat
84841  cagcagacca aagtaactta ctgtgtctga ttagctgttt ttaaattgtt gcttatattt
84901  taaaggcaaa aaaaaaacta agtgcttatt taacaacaac aaaaaatccc ctttacattt
84961  ttttagaggt attatttttct ctcacctatg aaactttctg accataaagt cagtgcgtaa
85021  tcactgatga ttcccagcat tctttcttta tccttgcttc ttttcatttt tcaattctta
85081  atccctcctt tcacttcccc cctgtattac aacaactaac agctgccaag aaagtggatt
85141  aggtcatata ttcaatcagg aggttttcct gttattcaag tcactttctt agtataccga
85201  aactaaaaaa tagttatggt ttttgcctca aaaaattcag ttggacttat attttattat
85261  ttgaaggttg ggtagcaaga agcaggtcta aacttacaaa agcaaagcag tcattttgga
85321  tcttcaactc cctatttctg tcttgagctg ttttcatttt gataatacat ctgaacactc
85381  ttctctaact tactgataat cagtccgtat tgttctacct cactttgtag acaatttgtc
85441  atcccatatc tacagctgca ggatcagtgg gacccatact ggtggagaaa accatgcagt
85501  ctctccatta ggaaaggaga atcagactga agactcctga cttgctttga tttctaataa
85561  taggctggtc tcactgggcc caggctgttc taaatctaga gtagttctga aatactgagt
85621  gtgaaatagg caatggtaga aattcttagg acccctgtaa gacaaatcaa aattccttct
85681  agtattcttt tctcccatca tttcctgcac tgYgctacgc caaatccaat ctcttaagga
85741  ccatttccta catgcaacca ttgagtgcac tttggcccca gtatagcttt ataaatctca
85801  ctagataact gtgggaagtc catccatgaa tcatagtatc aatggtctgc agctcagaaa
85861  gcacaaaaga acaaaataaa agtttgaaaa cttacagctt tctctgccag atatttttt
85921  gccaaatatc ttagactcct ttgtgttgct gtaacagaat gccacagact ggtaatttat
85981  aatgaacaga aatgtatttc gctcatggtt ctggagtctg gaagtcaaag aacatggcca
86041  gcatctgatg aggaccttca tgcagcatca ttccatggca gaagatagaa gagcaagaga
86101  ggatgatagc atgtgagaaa gtgccaaact tgattttata aaaaacccac ttccaggcca
86161  ggtgcaatgg ctcacacatg taatcccagc actttgggag gccacagcag gagaatcgct
86221  tgagcccaag agtttgagac caacctgggc aacataggga gatactgtct atacaaaaaa
86281  tgaaaaagtt agctaggagt ggtggcgcgt gccttgtagt cccagctact taggaggctg
86341  atgtgtgagg attgtttgag cctgagaggt caaggctgca gtgagctgtg attgcaccag
86401  tgtattccag cctgggcaac agagcaagac cctgcctctg aaaaaagaga aagaaaccca
86461  ctcccacagc attaattaat tcattcattc atgagggcag aggcctcatg acttaatcac
86521  cttctaaaag tcccatttct aaacactgtt gcattgggga ttaagtttct aacacataag
86581  ttttggagga cacattcaaa acatagcact aaatattcat tgggaactca tgtacttgga
86641  atatctactg gcaaaaaaaa aaaaaaaaga aagaaagaaa aaagaaaaaa attccatatt
86701  ccgtgttccc ataattgtgg tttatattta gtgaagcatc aaatgagcat gagatacaac
86761  tatttttttt atttacacaa aacttgaccc taaaatattt aaccaacaga agtagtacta
86821  ataaaattat tctatgaagt aatttttaat gaagctgagt ttattcaagt caYgtcttct
86881  gcaaataaaa atggacacca ataaacaaaa acaaagatag aaagaataac tgtgttcttg
86941  atatctcccc taaagttcac aatctctaca cctgtttctt tcatcttctg gtgatgttaa
87001  tcttctgttt attttgcgct ttaaatctaa gcacatgtgg attacccaga gattgccctc
87061  tgaaagtcag tctacacctg ttctttctta cctcacaaaa agtaatggaa aaaaaagtgt
87121  gtgtgtgtgt gtgtgcgtct gtgtgtgtgt gtgtcctgtt ggtggtagtg ttggtggtta
87181  aaaagcaatt tgggacttcc tctttgaaca gttgcctttt cctctcacag aaggaagatt
87241  tcatttYgtt tgagacgaga aaccaaacca cacaccaaag agaggggtat gatggctaag
87301  aagcccccaa aaccagcccc tcgcaggatc ttccaggaaa ggttaaagat tactgctcta
87361  cctttgtact ttgaaggttt tttattaatc aagcggtcag gataccgggt gagtctatag
87421  atgataatgt taaacctaag acttctgttt taatttaata tttatttcat ggtgatgtga
87481  tgtgttaaga cctccttgtt tctgttgaaa ttaaatcatc ttctcttctt tgagctcaga
87541  aaaatgattc caatttttca taatttaaat acaatgtctg gcttaaacct gtatgtatac
87601  acatatataa tatgtataat aacagaggtt gtaatattaa ggcactaata taaaaaagtc
87661  aataagctaa atttccaaag aatttattta aatatcacaa aagattttgg cttgggagat
87721  aaaatgtttc ttgtattttc tccacaattt attgctatgc ttcaatgatg acatgtacca
87781  ttaagataaa atgatatcat gattaaaatt aaacctgctc ctgttctgag tcatttgaag
87841  tttataatga tcaaattatt aaaaatggct tttgtaaaaa ttgttaaaat gacaaagttc
87901  atactgttta acattatata tagttatgtg ttctaaaata ctattcaaga tagtgacttt
87961  taatttttgg ggtactactg tgggtattaa gtacactaag ctacataaat acctttgctc
88021  taagaaaatc catgaagcat tctgtatttt taaatgtaat aattaaaact tatagttagt
88081  taaaatcatt acttttaaaa cagtaattat ggatgacttg aaattaatta gagaaataag
88141  cccaaaattg cctgttatta aataaaaaaa tcattaagtt aggtcaaatt ttatgaaatt
```

FIGURE 4-X

88201 gtatactgac taaaactaga aaaatttaa ggttttcaga aattccatca gaaatgttta
88261 atgatgctaa aatatatttt ctgaggattt atgaatactt gtggaaaaat tatatatatg
88321 aaaaatctat aatagcatat tcacatttct tacatatata atcagatcat ttactatttg
88381 agagtaaaga catggtaatt tgtattctgt tatggatgtt aaacatgcat aaataattac
88441 ctttcagtta tattagaatt ttttagattg atcctatatg cttttaatgt aaattcaatc
88501 ttgtcaccac aggtaagcca catagtcaca ctttgccagg aaagggaagt tgagaaaaaa
88561 aaattctaat tagtaattta aatcaggttg ttcattgaat gttttccaag gtatttataa
88621 taactgttta tgatagcagt ttttttaaa tgcttaaaga agacatgtca ttggtatggt
88681 tagcagaaag aataagaatt ttagagtgtc ctttatctga tggttttgcc acttatagct
88741 tgtgatcttg tgcaagttac tcaatctcct tgagactgtt tcaactataa gatagagagt
88801 atactacttg ctacttgcct cctaaggtac attctaggat tcagtaggtg ttcaatcaat
88861 ttattcactc aacatttatt gtgtgcctcc tatgcaccaa gcatcactca tttctggaga
88921 atatggaaaa aaacacaaag attttatttt caagaggctt aatatagtag gaatgatgtc
88981 ttccttacca aatttctact ctttaccttc tcttagaaag cattctttga agcagaatga
89041 atacctatag gcataaatat ttccaatgaa attaacttgt gtttctattt gaatttatag
89101 taaagtatct ttgtgtgtgt gggtgtgtgt ctgtgtgtgt gtgtgtgtgt gtgacagtgt
89161 ctcctctgtt gcccaagctg gagtgcaatg gtgcaatcac tagtcatgca gccttgacat
89221 ccccggctcc ggtgattctc ctgcctcagc ctcctgagta gctgggacta caagcgcata
89281 ccacacccgg ctaacgtttg tattgtttat agagacaggg tttcatcacg ttgttcaggt
89341 tggtctcaaa ctcctgggct caagcaacct gcccaccttg gcctcccaat gtgctagaat
89401 tacaggcatg ggcctctgag ttagaccact aagtaccttt tacttgtatg tcagggagaa
89461 gagagcaaga ggatgacaat aatacctact tatgtggtgg tttcagggat taaagggata
89521 gcatatgtaa aacacctggc tcacactaaa ggttagattc attctcttac cctttcatca
89581 cttatcatac tcttattcag gtactaaaat tagtttgagg tctgcaagta atatgactcc
89641 aaggagagtg agcatggtga taattagagt acttgaaaat agaagctatg agaaaaatct
89701 aagcaaaata agtggaattt ccaagcaatt ggcagcaaag tgccagggaa tctttgaaca
89761 gaaggtgagc caaaggtata tagccaagta atctttggag ctgattggct agaggaaggt
89821 gagaggctct gcagaatgta aagttgggtc tctgccaaag tcattcagaa catttaaact
89881 agaaaaaata taactgaaga gtaacctggt ctataaaggc atgaaatgct gagaggtgct
89941 tcttagctat tttttgtggg gcacttttaa gtatttaaag agcttttccc caggaaaatg
90001 taacaacgta taccattttg catgcaattc tagaaggttt ctaaatctga ggtcaagaac
90061 ccatgaaaaa cacataatct ttatggaaaa gataatgtgg gaagatttat ttaagccaca
90121 gcactaggat tctgaattaa ccaggaagat gtgtttatta tcaaagctgt taaatgccac
90181 agtggaatac cattggcatc gcagttctta caaatgggaa aatttctcat tttttctggg
90241 accgtttgaa aataagagag agtgtgggtg ggccgggagc ggtggctcat gcMtgtaatc
90301 ccagcactct gggaggccca ggtgggcgga tcacgaggtc aggaaatgga gaccaacctg
90361 ttcaacatgg cgaaaccccg tctctactaa aaatacaaaa attagcaggg catggtggca
90421 cgcacctgtg gtcccagcta ctcgggaggc tgaggcagga gaatcacttg aacctgggag
90481 tcagaggttg cagtgatccg agatcgcgcc actgcactcc agcctggcaa tagagcgaga
90541 ctctgtcaaa aaaaaaaaaa aaaaaaaaaa gaaagaaaga gagaaagaaa gaaaagaaaa
90601 agaaagaaaa taagagatag tttgggcaac gtctcataat ttcttcaatc tgtgatgtgc
90661 taaagtaaaa aaaaaaaaca gtgtgtgtgg gccgggcccg gtcccagcta ttcggaaggc
90721 tgaggcagga gaatcgcttg aatctgggag gcggaggttg cagtgagtca agatcgtgcc
90781 actgcactcc agcctgggtg acagagcgaa actccgtctc aaaaaaaaag gaaatttttt
90841 agggctgtct aagttgtgct cacacacagt tctctatggc agtaatctcc aaattttaaa
90901 ataacacact cctttcagga aaacattttg agcacagatc cctaatataa gaatattaat
90961 tcagccgggc gtggtggctc acgcctgtaa tcccagcact ttgggaggcc gaggcgggcg
91021 gatcacgagg tcagaagatt gagaccatcc tggctaacaa ggtgaaaccc tgtctctact
91081 aaaaaaaaaa aaaaaaaaaa attagccggg cgaggtggtg ggcgcctgta gtcccaacta
91141 ctcaggaggc tgaagcagga gaatggcgtg aacccaggag gcggagcttg cagggagcgg
91201 agattgcgcc attgcactcc aacctgggca acagagggag actccatctc aaaaacaaca
91261 aaaaaagaat attaattcat ttggaaatta taaatatata tattcccgta ctattaatcc
91321 atgtaaatta tttgatatat aaaaaatgaa ttaagaaata tgaaataaaa tataattaaa
91381 tattctaaaa ttttctctcc caatggatca tcttgcacac ctttggaatt tatgcatccc
91441 attttggaga ccactataca acattcttcc ttaggtaaat gtactttact tgcgcacaaa
91501 aacaaagaat gagagctttt tattggggac agaggaagag aaacttaact ggagcttgca
91561 ttgtaattca tttttgctag ttaattcatt tttgctagtt gtattaatga gaaaagagca
91621 ctgggaagtc aagagataaa agtatcagtc caattctaga atttgggcga ttccttcaca
91681 tttttaaatc tgttctctcc tttctaaaaa gaataattac tctcttgcat agctgaatag
91741 gttgctccaa tgactaagtg gactatataa taatgaatgc aacctttga ggtataacat
91801 ataatggatM ccattctagg ctcttagtgt tctgtcactg tgatccttac agtaactctt
91861 ctgggtagaa attattttat tcacttcaca aatgagtgag taggtgaggc atgctggctc
91921 atccctgtaa tcccagcact tatggaggcc aaggcaggag gattgcttga ggccaggagt
91981 tggagatcag cctaggcaat atagcgaaac cccatttcta caaaatcaaa aaattagcca

FIGURE 4-Y

```
92041   ggcgtggtgg cacaccctgt attctcaggt actcgggagg ctgaagtggg aggatcactt
92101   gaacccagga gttcgaggct gcagtgagct atgattgcat cactgcaccc cagcctgagc
92161   aacagagcga gaccctgtct ctctaaaaca acaactccaa ataaacaaat gagtgagttc
92221   tggccacaga ggttattcaa ggttcctctt ccattaagga ggctagcttg gatttgattg
92281   tgatgcactt gaccatgctg cttctcaaac tcaaattcta ccatttgatt gtaggtttcc
92341   tagttaagct agtgataaat tcagaagtct tatttcagcc tcttttctta catcaatgat
92401   ttcacacaaa cgattgttaa aaaaaaaaaa aaagaaattt gtgcttcatt ccaaagatat
92461   gtccttccaa attaattttt ttaattttta agtatttctc tagatatgaa taggtaccta
92521   gatgttgttt attgagattg atatttataa atgactcaga ttctatgtgt atttattaca
92581   tcttatcgaa atgaaatcaa gattgtgaYc tgtaagtctc tgcatactga atataagttt
92641   ggtctgtcct taggtttgca gctgtggttg cagttcacat ggtctaatgc tatttgtaac
92701   cttacttttc ctccagggct gcctttgctt tcaacttagt gcctttatac ttgcttttcc
92761   ccccatttaa taaatcctta atcctttcca caagtatata cgtttgctta cctgtaaaac
92821   tcccatttt ccctatttaa ctttacccaa ttcatatcat tcttttgaga ctgaacacct
92881   ctaYaatttt tctcgatagt gcaatggagg tctgcatgta catacagagg gaattcaata
92941   aacctttact ggctatcagt aatactagtt tttatacctt atggcagggt aatactgtag
```

FIGURE 5

NM_004087 [gi:4758161] Homo sapiens discs, large homolog 1 (Drosophila) (DLG1), mRNA

Gttggaaacggcactgctgagtgaggttgaggggtgtctcggtatgtgcgccttggatctggtgtaggcgaggtcac
gcctctcttcagacagcccgagccttcccggcctggcgcgtttagttcggaactgcgggacgccggtgggctagggc
aaggtgtgtgccctcttcctgattctggagaaaaatgccggtccggaagcaagatacccagagagcattgcaccttt
tggaggaatatcgttcaaaactaagccaaactgaagacagacagctcagaagttccatagaacgggttattaacata
tttcagagcaacctctttcaggctttaatagatattcaagaattttatgaagtgaccttactggataatccaaaatg
tatagatcgttcaaagccgtctgaaccaattcaacctgtgaatacttgggagatttccagccttccaagctctactg
tgacttcagagacactgccaagcagccttagccctagtgtagagaaatacaggtatcaggatgaagatacacctcct
caagagcatatttccccacaaatcacaaatgaagtgataggtccagaattggttcatgtctcagagaagaacttatc
agagattgagaatgtccatggatttgtttctcattctcatatttcaccaataaagccaacagaagctgttcttccct
ctcctcccactgtccctgtgatccctgtcctgccagtccctgctgagaatactgtcatcctacccaccataccacag
gcaaatcctcccccagtactggtcaacacagatagcttggaaacaccaacttacgttaatggcacagatgcagatta
tgaatatgaagaaatcacacttgaaaggggaaattcagggcttggtttcagcattgcaggaggtacggacaacccac
acattggagatgactcaagtattttcattaccaaaattatcacaggggagcagccgcccaagatggaagattgcgg
gtcaatgactgtatattac**[a/g]**agtaaatgaagtagatgttcgtgatgtaacacatagcaaagcagttgaagcgt
tgaaagaagcagggtctattgtacgcttgtatgtaaaaagaaggaaaccagtgtcagaaaaaataatggaaataaag
ctcattaaaggtcctaaaggtcttgggtttagcattgctggaggtgttggaaatcagcatattcctggggataatag
catctatgtaaccaaaataattgaaggaggtgcagcacataaggatggcaaacttcagattggagataaacttttag
cagtgaataacgtatgtttagaagaagttactcatgaagaagcagtaactgccttaaagaacacatctgattttgtt
tatttgaaagtggcaaaacccacaagtatgtatatgaatgatggctatgcaccacctgatatcaccaactcttcttc
tcagcctgttgataaccatgttagcccatcttccttcttgggccagacaccagcatctccagccagatactccccag
tttctaaagcagtacttggagatgatgaaattacaagggaacctagaaaagttgttcttcatcgtggctcaacgggc
cttggtttcaacattgtaggaggagaagatggagaaggaatatttatttcctttatcttagccggaggacctgctga
tctaagtggagagctcagaaaaggagatcgtattatatcggtaaacagtgttgacctcagagctgctagtcatgagc
aggcagcagctgcattgaaaaatgctggccaggctgtcacaattgttgcacaatatcgacctgaagaatacagtcgt
tttgaagctaaaatacatgatttacgggagcagatgatgaatagtagtattagttcagggtcaggttctcttcgaac
tagccagaagcgatccctctatgtcagagccctttttgattatgacaagactaaagacagtgggcttcccagtcagg
gactgaacttcaaatttggagatatcctccatgttattaatgcttctgatgatgaatggtggcaagccaggcaggtt
acaccagatggtgagagcgatgaggtcggagtgattcccagtaaacgcagagttgagaagaaagaacgagcccgatt
aaaaacagtgaaattcaattctaaaacgagagataaagggcagtcattcaatgacaagcgtaaaaagaacctctttt
cccgaaaattccccttctacaagaacaaggaccagagtgagcaggaaacaagtgatgctgaccagcatgtaacttct
aatgccagcgatagtgaaagtagttaccgtggtcaagaagaatacgtcttatcttatgaaccagtgaatcaacaaga
agttaattatactcgaccagtgatcatattgggacctatgaaagacaggataaatgatgacttgatctcagaatttc
ctgacaaatttggatcctgtgttcctcatacaactagaccaaaacgagattatgaggtagatggaagagattatcat
tttgtgacttcaagagagcagatggaaaaagatatccaggaacataaattcattgaagctggccagtataacaatca
tctatatggaacaagtgttcagtctgtacgagaagtagcaggaaagggcaaacactgtatccttgatgtgtctggaa
atgccataaagagattacagattgcacagctttaccctatctccatttttattaaacccaaatccatggaaaatatc
atggaaatgaataagcgtctaacagaagaacaagccagaaaaacatttgagagagccatgaaactggaacaggagtt
tactgaacatttcacagctattgtacaggggatacgctggaagacatttacaaccaagtgaaacagatcatagaag
aacaatctggttcttacatctgggttccggcaaaagaaaagctatgaaaactcatgtttctctgtttctcttttcca
caattccattttctttggcatctctttgccctttcctctggaaaaaa

FIGURE 6

NM_014660 [gi:7662303] Homo sapiens PHD finger protein 14 (PHF14), mRNA

ctgaggcagccgccctcgcgctgtgcaatttctggtctttcgttgcttctggtccaggctaataaagtttttctttc
tttaatttttttttcttctagttttaacgggagaaattaactccccggggccgccgggttgactgcgctgcctgggcc
ggaggtcttctccggccagggagcgctgtgggaaggggctcgagcggccagggccaggcgaggccgggggggcgggg
ggttaggggaccgcggggctactcttgggagcgcccctgtccggctggctgcgcgccggttttaaatagcatctttc
ggacttgtcttcgcggccccagtccccgacctcggcgctgcctgggctcctgcagcctctccctaagtcttctccaa
acgaccacctcacggattccttatggatcgcagctccaagaggaggcaggtgaagcctttggcagcttctctgctgg
aagctcttgattatgatagttcagatgacagtgatttaaagttggagatgcctcagattctgaagggagtggtaat
ggaagtgaagatgcttcaaaggacagtggagaaggttcctgtagtgattctgaagaaaatattttagaagaagaact
gaatgaagatattaaagtaaaagaagaacaacttaaaaattctgcagaggaagaagtactatcatcagaaaaacaat
taattaaaatggaaaagaaggaagaagaagaaaatggagaaagacctagaaagaaaagggagaaagagaaggaaaaa
gaaaaggaaaaggagaaagagaaggaaagagagaaggaaaaagaaaaagcaacagtatctgagaatgtggctgcttc
tgctgctgccaccacaccagccacaagtcctcctgctgttaacacatccccttctgttcccactacgacaaccgcta
cagaggaacaagtcagcgagccaaaaaaatggaaccttcgacgaaaccgaccacttctggattttgtgtccatggaa
gagctgaatgacatggatgactatgacagtgaggatgacaatgattggcgacctactgtagtaaagagaaaagggag
atctgcatctcagaaagagggaagtgatggagacaatgaggatgatgaagatgagggaagcgggagtgatgaagacg
agaatgatgaaggcaatgatgaagatcatagtagccctgccagtgaaggggttgcaagaagaagaagagtaaagtt
cttagcagaaacagtgctgatgatgaggaactgaccaatgatagcctgaccctatctcaaagcaagagtaatgagga
ctcgctgattcttgagaagagtcaaaactggagctctcaaaaaatggaccatattctgatttgctgtgtttgtctgg
gagataatagtgaggacgctgatgaaataattcagtgtgacaattgtggcattacagtccatgaaggttgttatgga
gttgatggagagagtgactctattatgagttcagcttctgaaaactccactgaaccttggttttgtgatgcctgtaa
atgtggtgtttctcctagctgtgaactgtgtcctaatcaggatggaattttcaaggagacagatgctggaagatggg
ttcatattgtttgtgccctgtatgttcctggagtagcctttggagatattgacaaattacgaccagtaacactaacg
gaaatgaactattccaaatatggtgccaaggagtgtagcttttgtgaagaccctcgctttgctagaactggggtttg
cattagctgtgatgcagggatgtgcagagcctatttccatgtgacctgtgctcaaaaggaaggtctgctttcagagg
cagcggcggaagaggatatagcagatccattctttgcttattgtaagcaacatgcagataggttagacagaaagtgg
aagagaaaaaactacttggctctacagtcctattgtaaaatgtctttgcaagagagagagaagcaactatcaccaga
agcacaggcaaggatcaatgcccggcttcagcagtatcgtgccaaagcagaactagctcgatctaccagaccccagg
cctgggttccaagggaaaaattgcccagaccactcaccagcagtgcttcagctattcgtaaacttatgcggaaagca
gaactcatggggatcagtacagatatctttccagtggacaattcagatactagttctagtgtggatggaaggagaaa
acataagcaaccagctctcactgcagattttgtgaattattattttgagagaaatatgcgcatgattcaaattcagg
aaaatatggctgaacaaaagaatataaaagataaattagagaatgaacaagaaaagcttcatgtagaatataataag
ctatgtgaatctttagaagaactacaaaacctgaatggaaaacttcgaagtgaaggacaaggaatatgggctttact
aggcagaatcacagggcagaagttgaatataccggcaattttgcgagcacccaaggagagaaaaccaagtaaaaaag
aaggaggcacacaaaagacatctactcttcctgcagtactttatagttgtgggatttgtaagaagaaccatgatcag
catcttcttttattgtgtgatacctgtaaactacattaccatcttggatgtctggatcctcctcttacaaggatgcc
aagaaagaccaaaaacagttattggcagtgctcggaatgtgaccaggcagggagcagtgacatggaagcagatatgg
ccatggaaaccctaccagatggaaccaaacgatcaaggaggcagattaaggaaccagtgaaatttgttccacaggat
gtgccaccagaacccaagaagattccgataagaaacacgagaaccagaggacgaaaacgaagcttcgttcctgagga
agaaaaacatgaggaaagagttcctagagagagaagacaaagacagtctgtgttgcaaaagaagcccaaggctgaag
atttaagaactgaatgtgcaacttgcaagggaactggagacaatgaaaatcttgtcagataccttcatgagaccca
actctgccacagctcatcctcggaggcaatcccggaaacctcttttataagtgtgattttgtcactgcaaaaagttgcct
tctcgatagagtacataaagtaaaggagaacagctatattgtcctttctataagcttgtcactgcaaaaagttgcct
tttgcttgtcaggtttggatagaatataattgattggtttgttacttggactaacaaggcagtagatttgcctgtgt
ggaatttttttcctatttttcttcttttgctttttttgcacttatcagaaatatttgatgtgtgcattgttgaaa
aatactggatacatgtcttgtcagaattgtcgtattaagtaatgtctttcccctcttgcttctttggcaaatgttgat
acacagtgtttggacttactgaagagttactgaagcctgtgggacttaataaatacaatatgctactccctgttgtg
attgaacaatacaaataacaaacaagaaagacaatgcattccattagtctgctattctgtttccttcaacttcatac
atagattcatatatgaagtactgcattgtaaaacaactatagactcatattaaatgttatttctatttataatattc
agcaaaagtgaaagacttgtgaagcatatgacattctatttttgacttattagtcctagtgtgaaagcattaatatt
attagcatgaaatttttacttcagattttaagctcatgaataaagatatatctgtgttgatctctagatatttttag
taatacccaaatatattcagtccttattgttttaatataagctttctgtgttcagtataatttttatttttctcaacc
ttaaatatgaacttaggaaataaggagggaagtacaaagattattgactatacaacctaccagctgaaagaaagatc
ttcatcaacatctgtatctttccagaggtatacagaattaaaatttgatcttcaagctttaatgatccagttttaag
tcaacggcagaagtatgttgaatatttcatcactcaatcttgaactgatttagaagagactctttgctgaaattgaa
ttgcacttatacatgtaaattgtcaacatgtaatttggaattttctgattaataaatgtggttttggacatct

FIGURE 7

NM_012074 [gi:13442997] Homo sapiens D4, zinc and double PHD fingers, family 3 (DPF3), mRNA

acattgtagcaaaatggcgactgtcattcacaacccccctgaaagcgctcggggaccagttctacaaggaagccattc
agcactgccggagttacaactcacggctgagtgcagagcgcagcgtgcgtcttcccttcctggactcacagactggc
gtggcccagaacaactgctacatctggatggagaagaggcaccgaggcccaggccttgccccgggccagctgtatac
ataccctgcccgctgctggcgcaagaagagacgattgcacccacctgaagatccaaaactgcggctgctggagataa
aacctgaagtggagcttcccctgaagaaggatgggttcacctcagagagcaccacgctggaagccttgctccgtggc
gaggggttgagaagaaggtggatgccagggaggaggaaagcatccaggaaatacagagggttttggaaaatgatga
aaatgtagaagaagggaatgaagaagaggatttggaagaggatattcccaagcgaaaggacaggactagaggacggc
ctcgctgccctctcccttccctgcactgtttttcctcccttcccctctgccgtgatagatgctaaggagtggggtgga
ggtgggaagtgggaagcaacagtggcgtataggaaaaagaaaatataccccgtgcacattttcaacatgtagttgaa
gaagcctaaattaggtactagaaaaaaaaaaaggacagaaacactgcctgatatgtgagcaagagcatgaaaatata
gaatgattatctcagagcagaggaggtggggagcaggtggctggagagaggcggggtaggtaggcagggcgtccctg
gctcaccaatgcgtctggtccacatggcatttgggaaagcagatgggtcatgcttttgcaatgaatggtcagatgct
taggggcatttgtgctcctctggccaggaaagggaacagagtccatccaagctgccccccactcatctgccatggt
tgtccaccctgggggtttttcctctgaaggaaaatgaacccatcttttgcctgccatgaatcattgcagggcaggct
gttactggctaaattaaaaccccttggcagggcatcgactgtgtgcgaggcaatgctctaggtcctgtaggggctgga
aaataaaccgcatgctggcctagccttcaagttgcttttcagccatgaaactaagctctgccaagcaatcgtgattg
taggtcaagtatcgcgccatcacggaaggtgtaagtatatgtaggtttctctgttgaggatgcgtgttcctcagtag
aagacaagcaggtggaggcatccagtgatttctaccctgtggaggctgaggggtcgggggaagaaaacagatgacct
cagcctagttgctttaatctgcttttccaagcactggatgcccttggatgacagcatcctcagcattaaaactggtg
aactgatgaagtcacctggcctggagtgtgttgggcagccagtgtccccagagctgcttgtgggtttctggggtgga
aggcagggaggtgcaactggcagggcctgatcagaggcagaaaatgacccccacagtggtcttttccctgctagaga
aagcagagagcgggactgggggggtgggggcttcaagtacagattgggcacactccaccagaccccagcaaggtcag
ctgccccacgcctgtatctggcactgctggtgtgtgcagggatgaaacccagcatcagagaggttttcagcaaacct
tccatgctcacctttggccaggtgcttgtcagatcctagcttcgtgctggcttagactctcagttgtttgtgttgac
agcattaggaataaaccgtttgtttcatctttccttttctccacacgtgtcacagccagatttcagctttgagttat
tccctgaagaagccacaccaaccttgctttcaagcaaaatgcctgggcttgggggaaggtgtgtatctgtccatgtg
tggatgtcggctcagagctatagcttctctgtgggggggtggcccaagggaaggctcctctggggccctggatggcac
atgactcccagtgaggagaattcaggtgatctctgtggaggtagtcagggacacaaggcttggctgtgagtctggtt
ttaaagtgcgtgacagcctgaaaagcatgcaggggtttggtccactcacctacttgaaagcctgtgggcaacgttct
tttgagccaagacttctctgaatggccctgctggtggaaggggtgaggcaaaggcctctgacttggacccttccac
accagactggcagcacttcccccaggcagccagtggtgggccctgagccctcaggtccccagctccttgagggatga
acctgggagcccaagagccagtggctgagctctgagaaggctccatctcccacctgcccttgagcgcgctctcaggc
tgagaacacggtctcatcaggcgccttcctggcctgatgctgtgctgtctacgtcacacggtcgattcacaaaagcc
agaactagacctcaaccaggtcatctccccgttgccaagtgggttcagggtgagggcaatttgtaagctcaatttct
ctgacagccaagacatggagcatctctgctaagaagccaaaagaaattggttttcttcttctcattgctgaagcccc
tctgtgtctcttctcagggacaggctggtccagtggctttggtgagggcgcctccatttgtgaacgctgggattcct
ctgggtgggctttggcaggtggctcctatggtaggaaggatgaccaggtccctgggaatgagctgcctacctgctgc
tccacgaggaagctcaggcctgcacaggctccaccaggccttggatgccctctagttgagtcagagaccctggaaac
acactgagatctccaattgctgcctccattgatgtctctagacctgcagatacgaagcaaacctgggattgcttctt
ccaggtatgggcaccagagagggaggccactgcaacattttatccccatcattccaaaatgcttgcttgtctctttt
acctcacactcacaactccttctgagactctcagtcataaaggaatgaccaagagagtgggtctccagtgagagaaa
tgcctatgaaagagggtttcccttttgctctttgaacaccctccccactgatccttgggacccaacgccgcattg
cctcttgcagatgaggttttgccttgggctgcttgggtacttcagaccaggactgagtctgacacagctttcatgag
gttacagaaaagggctacagatttgggaagctgtgtgtaatggtcttgagacaatatctccatttggcccaccctgg
cttctctaaaaagcaacgacagcaacagacaaacaaaaagctcccacctcccaccccgttagctgtcctcctccttc
actgtgatgtggttgcggtctctgtaggtgtgtgtgccacccttgtcctctgtcctctggggatgtgcccttcccac
gtgtgtcaggttcccactctttcgtggttcctaacgtgaagtgctgtgatgtttctgccctgcctaaggaacgtatc
aagctctctcagtgtttcagtgttggagattgaggctgtgccacatcttctgccatcctaaggggacatgatggttc
tgtgattcccagagagctggcagattgtgacaatctccaggagaacctacagattggaagcagcccacacctgatgt
ggactcctgtcccgggactcactcttcattcagaagactggtggcccacgtgccaggaccaccccacctctcttgct
gccttttctcctgtcctgatggggttctgggagggagacctgtcgctgatgagatgaagaatgtggggatcgagcag
ccttcttctttgggacccctcgatatcccatggaatgctcgcacgttctcaaagactgagtcacaagcccctacccc
ttccttgctgtggttagtatcttgttctgtgattggttagcaatgttgactacccacgtagtgaatcttttgtctgc
aatttagagaatgtgtaaacaaataaaaggctttaaaactc

FIGURE 8

NM_001812 [gi:4502778] Homo sapiens centromere protein C 1 (CENPC1), mRNA

cggatcgcagctctcgcggcagtcgcctgagacttaaggttattgcttggccgcggcctggtattccggcgattcgt
ttcttgctcggcttcctggagctgtggtccgtgtgggcttccacctcagacagttgcgctggctcagcggggccgga
acatggctgcgtccggtctggatcatctcaaaaatggctacagaagaagattttgtcgaccttccagggcacgtgac
attaacacagagcaaggccagaatgttctggaaatcttacaagactgttttgaagaaaaaagtcttgccaatgattt
tagtacaaattctacaaaatcagtgcctaattcaacacgcaaaataaaagacacttgtattcagtcaccaagcaaac
agtgccagaaatcacatccaaagtcagttccagtttcttcaaagaagaaagaagcctctctacagtttgttgtagaa
ccaagtgaagccacaaacagatcagttcaggcccatgaagttcatcagaaaattctggcaactgatgttagttccaa
aaatacacctgactcgaaaaaaatatcaagtagaaacataaatgatcatcacagtgaagctgatgaagaattttact
tatccgttggctcaccttctgttcttttggatgcaaaaacatctgtatcacaaaatgttattccatctagtgccaaa
aagagagagacttacacttttgaaaattcagtaaatatgctgccttcaagtacagaggtttcagttaaaaccaaaaa
aaggttaaactttgatgataaagttatgttaaagaaaatagaaatagataataaagtatcagatgaagaggataaaa
catcggaaggacaagaaagaaaaccatcaggatcatctcagaatagaatacgagattcagaatatgaaattcaacga
caagctaaaaaaagtttttcaacattgtttttagaaacagtaaaacgaaaaagtgaatccagtcccattgttaggca
tgcggcaactgctccacctcattcgtgtcctcccgatgatacgaagttgatagaggatgaatttataattgatgagt
cggatcaaagttttgccagtagatcttggattacaataccaagaaaggcagggtctctgaaacaacgcacaatatcc
ccggctgagagcactgcactctttcaaggtagaaagtcaagagaaaagcatcataatatattacctaagactttggc
aaatgacaaacattcccataaacctcacccagtagagacatctcagccctctgataaaacagtactggatacaagtt
atgctttgatagatgaaacagtaaataattatagatctacaaaatatgaaatgtattccaagaatgcagaaaaacca
tctagaagcaaaaggactataaaacaaaaacagagaagaaaattcatggctaaaccagctgaagaacagcttgatgt
gggacagtctaaagatgaaaacatacatacatcacatattacccaagacgaatttcaaagaaattcagacagaaata
tggaagagcatgaagagatgggaaatgattgtgtttccaaaaaacagatgccacctgtgggaagcaagaaaagtagc
actagaaaagataaggaagaatctaaaaagaagcgcttttccagtgagtccaagaacaaacttgtacctgaagaagt
gacttcaactgtcacgaaaagtcgaagaatttccaggcgtccatctgattggtgggtggtaaaatcagaggagagtc
ctgtttatagcaattcttcagtaagaaatgaattaccaatgcatcacaatagtagccgaaaatctactaagaaaaca
aatcagtcatctaagaatattaggaaaaaaactattccacttaaaaggcagaagacagcaactaaaggcaaccaaag
agtacagaagtttttaaatgctgaaggttctggaggtatcgttggtcatgatgaaatttccagatgttcactgagtg
agccattggaaagtgatgaggcagacttggctaagaagaaaaatcttgattgttctagatctacaagaagctcaaag
aatgaagataacattatgactgcacagaatgttcccctaaagcctcagaccagtggatatacatgtaatataccaac
agagtcaaacttggattctggagagcataagacttcagttttagaggaaagtggaccttccaggctcaataataatt
atttaatgtctggaaagaatgatgtggatgatgaggaagttcatggaagttcagatgactcaaaacaatctaaagtg
ataccaaagaacagaatccatcacaaactagtattgccctccaacacaccaaatgttcgcaggaccaagagaacacg
tttgaaacctttggagtactggcgaggagagcgaatagattatcaaggaaggccatcaggaggattcgtgattagtg
gagtactatctccagacacaatatcgtctaaaaggaaggcaaaagaaaatattggaaaagtcaacaaaaaatctaat
aagaaaaggatctgtcttgataacgatgaaagaaagactaacttaatggtaaatctaggtatacctcttggagatcc
tttgcagccaacgagggtaaaggacccagaaacaagagagattattctcatggatcttgtaaggccacaagatacat
atcaatttttgttaagcatggtgagttgaaggtatacaagacattggatacacccttttttctactgggaaattg
atattaggaccacaagaagaaaagggaaagcagcatgttggccaggatatattggtttttatgttaactttggtga
ccttttgtgtactttacatgaaacaccttatatattaagtactggggattcgttctatgttccttcaggtaactatt
ataacatcaaaaatctccggaatgaggaaagtgttcttcttttactcagataaaaagatgaaagatcaaccaacct
taaatatatgtatgtatatatgtatatgtaaaaacagtttgtatagttggaatatttgtctttgtaattacttgtga
tgttttaaaataaaaatttttattcagttttgtgtaaaaaaaaaaaaaaaaaa

FIGURE 9

NP_004078 [gi:4758162] synapse-associated protein 97; discs large homolog 1; presynaptic protein SAP97

MPVRKQDTQRALHLLEEYRSKLSQTEDRQLRSSIERVINIFQSNLFQALIDIQEF
YEVTLLDNPKCIDRSKPSEPIQPVNTWEISSLPSSTVTSETLPSSLSPSVEKYRY
QDEDTPPQEHISPQITNEVIGPELVHVSEKNLSEIENVHGFVSHSHISPIKPTEAV
LPSPPTVPVIPVLPVPAENTVILPTIPQANPPPVLVNTDSLETPTYVNGTDADYE
YEEITLERGNSGLGFSIAGGTDNPHIGDDSSIFITKIITGGAAAQDGRLRVNDCI
LQVNEVDVRDVTHSKAVEALKEAGSIVRLYVKRRKPVSEKIMEIKLIKGPKG
LGFSIAGGVGNQHIPGDNSIYVTKIIEGGAAHKDGKLQIGDKLLAVNNVCLEE
VTHEEAVTALKNTSDFVYLKVAKPTSMYMNDGYAPPDITNSSSQPVDNHVSP
SSFLGQTPASPARYSPVSKAVLGDDEITREPRKVVLHRGSTGLGFNIVGGEDG
EGIFISFILAGGPADLSGELRKGDRIISVNSVDLRAASHEQAAAALKNAGQAVT
IVAQYRPEEYSRFEAKIHDLREQMMNSSISSGSGSLRTSQKRSLYVRALFDYD
KTKDSGLPSQGLNFKFGDILHVINASDDEWWQARQVTPDGESDEVGVIPSKR
RVEKKERARLKTVKFNSKTRDKGQSFNDKRKKNLFSRKFPFYKNKDQSEQET
SDADQHVTSNASDSESSYRGQEEYVLSYEPVNQQEVNYTRPVIILGPMKDRIN
DDLISEFPDKFGSCVPHTTRPKRDYEVDGRDYHFVTSREQMEKDIQEHKFIEA
GQYNNHLYGTSVQSVREVAGKGKHCILDVSGNAIKRLQIAQLYPISIFIKPKS
MENIMEMNKRLTEEQARKTFERAMKLEQEFTEHFTAIVQGDTLEDIYNQVKQ
IIEEQSGSYIWVPAKEKL

DLG1 Domains

Gene	Prediction Method	Accession ID	Domain Description	Start	End
DLG1	Pfam	PF00595	PDZ	218	304
DLG1	Pfam	PF00595	PDZ	313	399
DLG1	Pfam	PF00595	PDZ	460	540
DLG1	Pfam	PF00018	SH3	578	644
DLG1	Pfam	PF00625	Guanylate_kin	741	843
DLG1	prosite	PS00856	Guanylate_kin	740	757
DLG1	pfscan	PS50106	PDZ	218	256

FIGURE 10

NP_055475 [gi:7662304] PHD finger protein 14 [Homo sapiens]

MDRSSKRRQVKPLAASLLEALDYDSSDDSDFKVGDASDSEGSGNGSEDASKD
SGEGSCSDSEENILEEELNEDIKVKEEQLKNSAEEEVLSSEKQLIKMEKKEEEE
NGERPRKKREKEKEKEKEKEKEKEREKEKEKATVSENVAASAAATTPATSPP
AVNTSPSVPTTTTATEEQVSEPKKWNLRRNRPLLDFVSMEELNDMDDYDSED
DNDWRPTVVKRKGRSASQKEGSDGDNEDDEDEGSGSDEDENDEGNDEDHSS
PASEGGCKKKKSKVLSRNSADDEELTNDSLTLSQSKSNEDSLILEKSQNWSSQ
KMDHILICCVCLGDNSEDADEIIQCDNCGITVHEGCYGVDGESDSIMSSASENS
TEPWFCDACKCGVSPSCELCPNQDGIFKETDAGRWVHIVCALYVPGVAFGDI
DKLRPVTLTEMNYSKYGAKECSFCEDPRFARTGVCISCDAGMCRAYFHVTCA
QKEGLLSEAAAEEDIADPFFAYCKQHADRLDRKWKRKNYLALQSYCKMSLQ
EREKQLSPEAQARINARLQQYRAKAELARSTRPQAWVPREKLPRPLTSSASAI
RKLMRKAELMGISTDIFPVDNSDTSSSVDGRRKHKQPALTADFVNYYFERNM
RMIQIQENMAEQKNIKDKLENEQEKLHVEYNKLCESLEELQNLNGKLRSEGQ
GIWALLGRITGQKLNIPAILRAPKERKPSKKEGGTQKTSTLPAVLYSCGICKKN
HDQHLLLLCDTCKLHYHLGCLDPPLTRMPRKTKNSYWQCSECDQAGSSDME
ADMAMETLPDGTKRSRRQIKEPVKFVPQDVPPEPKKIPIRNTRTRGRKRSFVP
EEEKHEERVPRERRQRQSVLQKKPKAEDLRTECATCKGTGDNENLVRYPS

FIGURE 11

NP_036206 [gi:13442998] cer-d4 (mouse) homolog; 2810403B03Rik [Homo sapiens].

MATVIHNPLKALGDQFYKEAIEHCRSYNSRLSAERSVRLPFLDSQTGVAQNN
CYIWMEKRHRGPGLAPGQLYTYPARCWRKKRRLHPPEDPKLRLLEIKPEVEL
PLKKDGFTSESTTLEALLRGEGVEKKVDAREEESIQEIQRVLENDENVEEGNE
EEDLEEDIPKRKDRTRGRARCPLPSLHCFSSLPSAVIDAKEWGGGGKWEATV
AYRKKKIYPVHIFNM

FIGURE 12

NP_001803 [gi:4502779] centromere protein C 1; Centromere autoantigen C1 [Homo sapiens]

MAASGLDHLKNGYRRRFCRPSRARDINTEQGQNVLEILQDCFEEKSLANDFS
TNSTKSVPNSTRKIKDTCIQSPSKECQKSHPKSVPVSSKKKEASLQFVVEPSEA
TNRSVQAHEVHQKILATDVSSKNTPDSKKISSRNINDHHSEADEEFYLSVGSPS
VLLDAKTSVSQNVIPSSAKKRETYTFENSVNMLPSSTEVSVKTKKRLNFDDKV
MLKKIEIDNKVSDEEDKTSEGQERKPSGSSQNRIRDSEYEIQRQAKKSFSTLFL
ETVKRKSESSPIVRHAATAPPHSCPPDDTKLIEDEFIIDESDQSFASRSWITIPRK
AGSLKQRTISPAESTALFQGRKSREKHHNILPKTLANDKHSHKPHPVETSQPS
DKTVLDTSYALIDETVNNYRSTKYEMYSKNAEKPSRSKRTIKQKQRRKFMAK
PAEEQLDVGQSKDENIHTSHITQDEFQRNSDRNMEEHEEMGNDCVSKKQMPP
VGSKKSSTRKDKEESKKKRFSSESKNKLVPEEVTSTVTKSRRISRRPSDWWVV
KSEESPVYSNSSVRNELPMHHNSSRKSTKKTNQSSKNIRKKTIPLKRQKTATK
GNQRVQKFLNAEGSGGIVGHDEISRCSLSEPLESDEADLAKKKNLDCSRSTRS
SKNEDNIMTAQNVPLKPQTSGYTCNIPTESNLDSGEHKTSVLEESGPSRLNNN
YLMSGKNDVDDEEVHGSSDDSKQSKVIPKNRIHHKLVLPSNTPNVRRTKRTR
LKPLEYWRGERIDYQGRPSGGFVISGVLSPDTISSKRKAKENIGKVNKKSNKK
RICLDNDERKTNLMVNLGIPLGDPLQPTRVKDPETREIILMDLVRPQDTYQFF
VKHGELKVYKTLDTPFFSTGKLILGPQEEKGKQHVGQDILVFYVNFGDLLCTL
HETPYILSTGDSFYVPSGNYYNIKNLRNEESVLLFTQIKR

FIGURE 13

DLG1

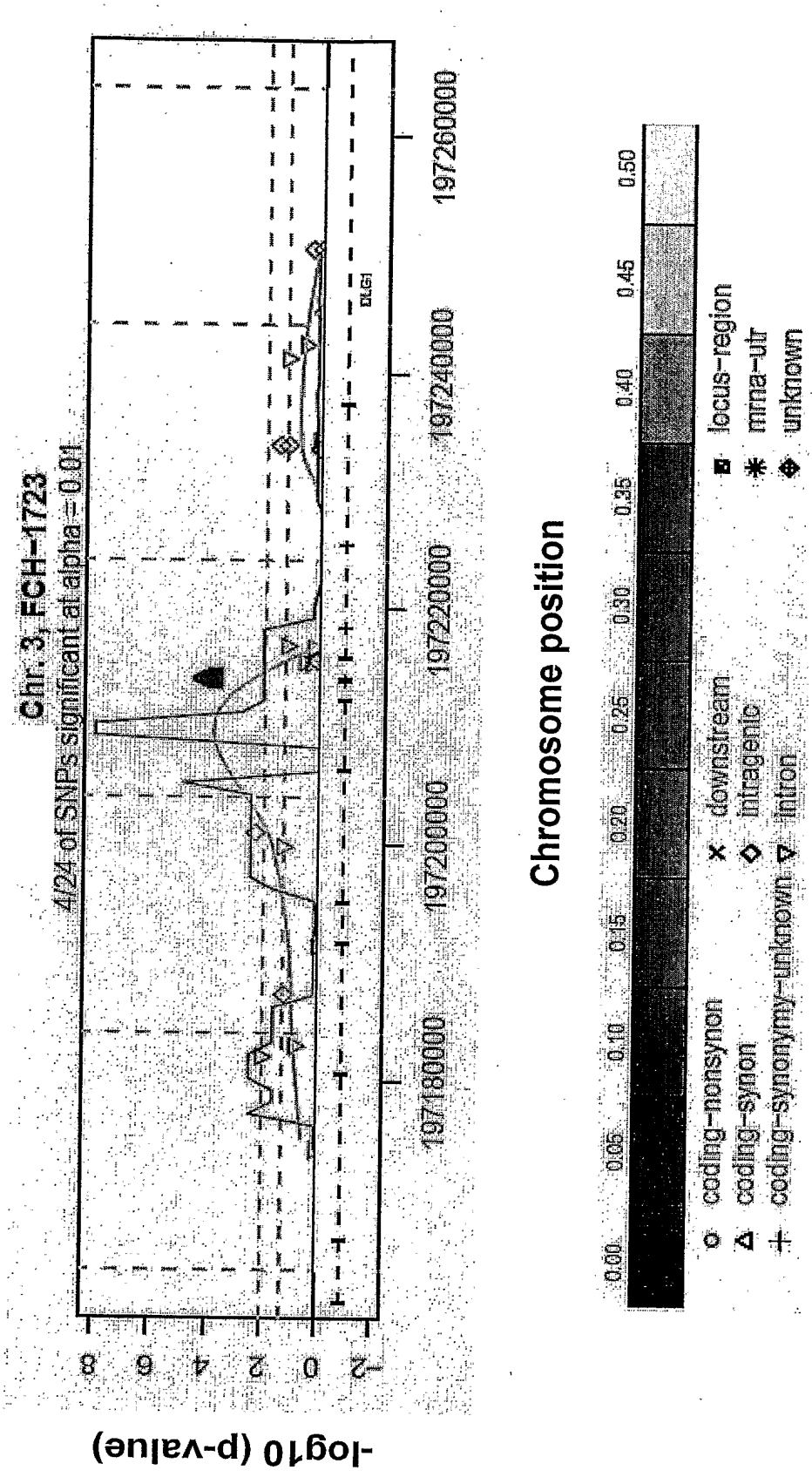


FIGURE 14

KIAA0783

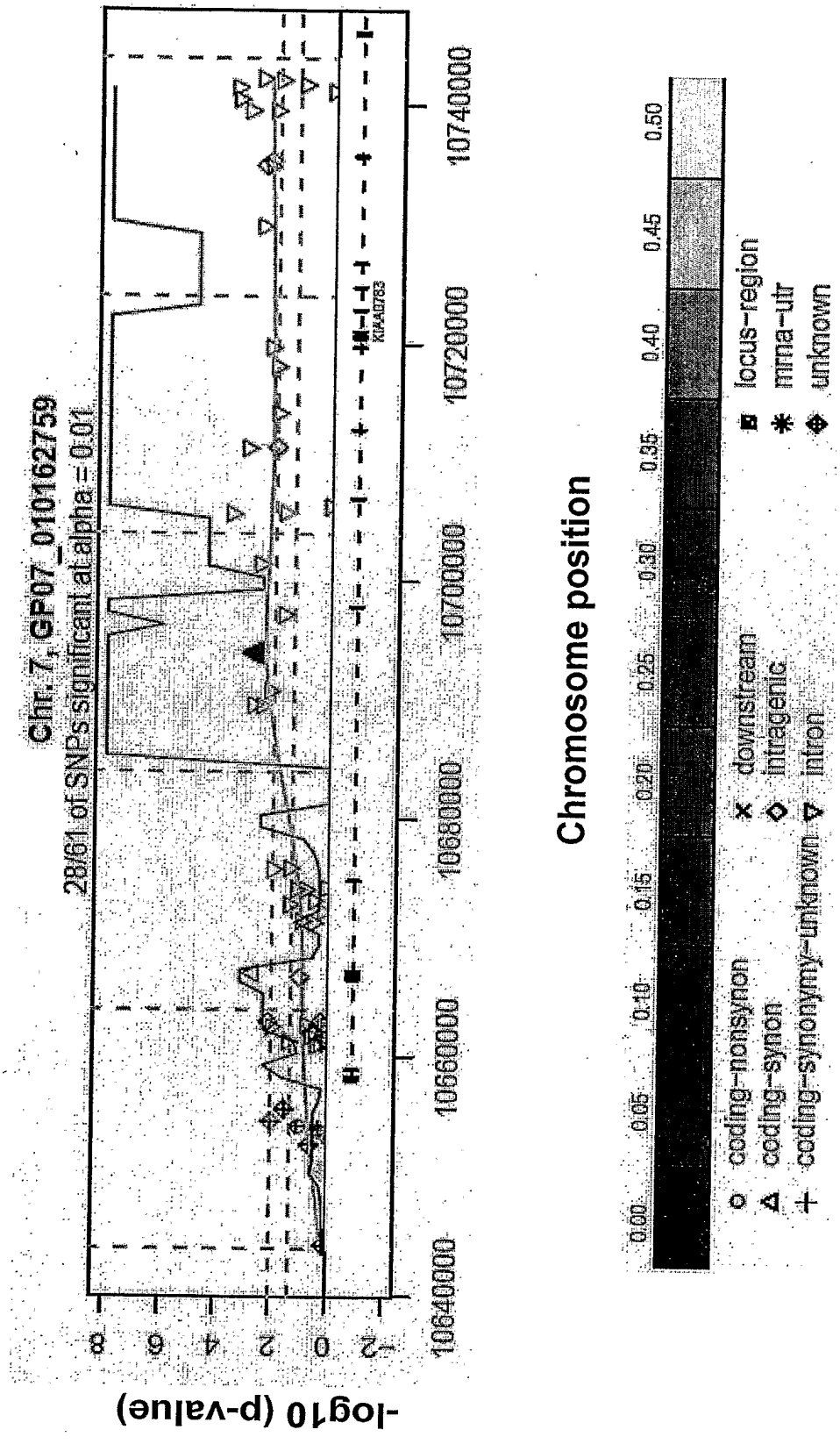


FIGURE 15

DPF3

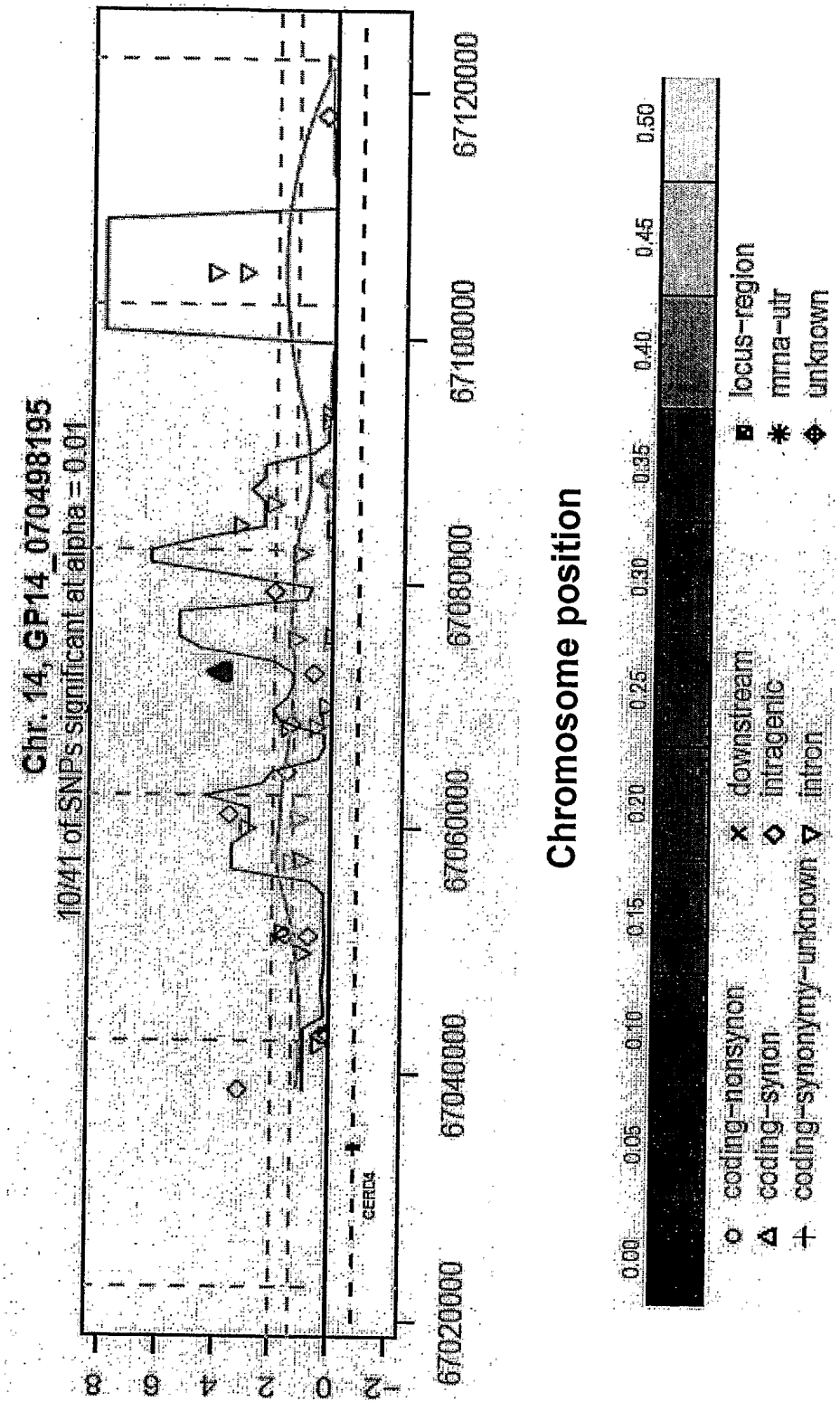


FIGURE 16

CENPC1